

DNA-BASED MARKERS IN PLANTS

Advances in Cellular and Molecular Biology of Plants

VOLUME 1

Editor-in-Chief

Indra K. Vasil, *Laboratory of Plant Cell and Molecular Biology,
University of Florida, Gainesville, Florida, USA*

Editorial Advisory Board

Robert T. Fraley, *St. Louis, Missouri, USA*

Robert B. Goldberg, *Los Angeles, California, USA*

Charles S. Levings, III, *Raleigh, North Carolina, USA*

Ronald L. Phillips, *St. Paul, Minnesota, USA*

Jeff Schell, *Cologne, Germany*

The titles published in this series are listed at the end of this volume.

DNA-based markers in plants

Edited by

RONALD L. PHILLIPS

*Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul,
Minnesota, U.S.A.*

and

INDRA K. VASIL

*Laboratory of Plant Cell and Molecular Biology, University of Florida, Gainesville,
Florida, U.S.A.*



SPRINGER-SCIENCE+BUSINESS MEDIA, B.V.

Library of Congress Cataloging-in-Publication Data

DNA-based markers in plants / edited by Ronald L. Phillips and Indra K. Vasil.

p. cm. -- (Advances in cellular and molecular biology of plants, v. 1)

Includes index.

ISBN 978-94-010-4482-0 ISBN 978-94-011-1104-1 (eBook)

DOI 10.1007/978-94-011-1104-1

1. Plant genome mapping. 2. Genetic polymorphisms. 3. Plant molecular genetics. I. Phillips, Ronald L. II. Vasil, I. K. III. Series.

QK981.45.D53 1994

581.87'3282--dc20

93-49927

ISBN 978-94-010-4482-0

printed on acid-free paper

All rights reserved

© 1994 Springer Science+Business Media Dordrecht

Originally published by Kluwer Academic Publishers in 1994

Softcover reprint of the hardcover 1st edition in 1994

No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the copyright owners

Contents

General preface	vii
Preface	ix
1. Some concepts and new methods for molecular mapping in plants B. Burr	1
2. RFLP technology G. Kochert	8
3. Constructing a plant genetic linkage map with DNA markers N.D. Young	39
4. Mapping quantitative trait loci S.J. Knapp	58
5. Breeding multigenic traits C.W. Stuber	97
6. Nuclear DNA markers in systematics and evolution R. Whitkus, J. Doebley and J.F. Wendel	116
7. Introduction: molecular marker maps of major crop species R.L. Phillips and I.K. Vasil	142
8. Molecular maps of alfalfa E.C. Brummer, C.S. Echt, T.J. McCoy, K.K. Kidwell, T.C. Osborn, G.B. Kiss, G. Csanadi, K. Kalman, J. Györgyey, L. Ökrész, A.E. Raczkevly, J.H. Bouton and G. Kochert	144

vi Contents

9. An integrated RFLP map of <i>Arabidopsis thaliana</i> H.M. Goodman, S. Hanley, S. Cartinhour, J.M. Cherry, B. Hauge, E. Meyerowitz, L. Medrano, S. Kempin, P. Stamm and M. Koornneef	159
10. RFLP maps of barley A. Kleinhofs and A. Kilian	163
11. DNA-based marker maps of <i>Brassica</i> C.F. Quiros, J. Hu and M.J. Truco	199
12. Genetic mapping in lettuce R.W. Michelmore, R.V. Kesseli and E.J. Ryder	223
13. RFLP maps of maize E.H. Coe and J.M. Gardiner	240
14. RFLP map of peanut T. Halward, H.T. Stalker and G. Kochert	246
15. <i>Phaseolus vulgaris</i> : the common bean C.E. Vallejos	261
16. RFLP map of the potato C. Gebhardt, E. Ritter and F. Salamini	271
17. Rice molecular map S.D. Tanksley	286
18. Generation of a genetic map for <i>Sorghum bicolor</i> J.L. Bennetzen and A. Melake-Berhan	291
19. RFLP map of soybean R.C. Shoemaker	299
20. Tomato molecular map S.D. Tanksley	310
21. RFLP maps of bread wheat G.E. Hart	327
22. Development of RFLP-based linkage maps in diploid and hexaploid oat (<i>Avena</i> sp.) L.S. O'Donoghue, P.J. Rayapati, S.F. Kianian, M.E. Sorrells, S.D. Tanksley, M. Lee, H.W. Rines and R.L. Phillips	359
List of contributors	375
Subject index	380

General preface

The double helix architecture of DNA was elucidated in 1953. Twenty years later, in 1973, the discovery of restriction enzymes helped to create recombinant DNA molecules *in vitro*. The implications of these powerful and novel methods of molecular biology, and their potential in the genetic manipulation and improvement of microbes, plants and animals, became increasingly evident, and led to the birth of modern biotechnology. The first transgenic plants in which a bacterial gene had been stably integrated were produced in 1983, and by 1993 transgenic plants had been produced in all major crop species, including the cereals and the legumes. These remarkable achievements have resulted in the production of crops that are resistant to potent but environmentally safe herbicides, or to viral pathogens and insect pests. In other instances genes have been introduced that delay fruit ripening, or increase starch content, or cause male sterility. Most of these manipulations are based on the introduction of a single gene – generally of bacterial origin – that regulates an important monogenic trait, into the crop of choice. Many of the engineered crops are now under field trials and are expected to be commercially produced within the next few years.

The early successes in plant biotechnology led to the realization that further molecular improvement of plants will require a thorough understanding of the molecular basis of plant development, and the identification and characterization of genes that regulate agronomically important multigenic traits. During the past ten years there has been a resurgence of molecular and related cellular studies in plants, including the molecular mapping of plant genomes. A great deal of interesting and useful information has been generated about the molecular basis of important plant processes. This series of volumes is intended to chronicle the most important advances in the cellular and molecular biology of plants, and to stimulate further interest and research in the plant sciences. The success and usefulness of these volumes depends on the timeliness of the subjects discussed, and the authoritative and insightful accounts provided by distinguished and internationally respected contributing authors. In this, I have been greatly aided by the advice of members of our Editorial Advisory Board and the editors of individual volumes, to whom I owe a debt of gratitude. I also

thank Dr. Ad. C. Plaizier of Kluwer Academic Publishers in helping me to launch this series, and his competent and helpful staff in the preparation of the volumes for publication. The various volumes already in press and in preparation have provided me the opportunity to know and work with many colleagues, and have helped me to improve my own understanding and appreciation of plant molecular biology.

Indra K. Vasil

Preface

As with numerous other technological advances, many insights have emerged from application of the new tools of molecular biology. The concept of DNA-based markers has revolutionized our ability to follow chromosome segments, including minute regions, and has led to new opportunities such as map-based cloning and directed plant breeding. Species with little genetic information available in the past now have hundreds of genetic markers. In some cases, the map from one species can be transferred almost directly to another species, such as from tomato to potato. Additional genomic structure information, such as homoeologous chromosome identification, is forthcoming for several species. The suggestion of ancient polyploidy has been made even for maize, the most genetically studied plant species. The ancestral relationships of species and the pedigree relatedness of lines have been identified in hundreds of situations.

This volume records the recently developed linkage maps of many important crop species. This recording, as impressive as it appears, will perhaps serve as a historical log of the development of the first molecular genetic marker maps for several species. The initial chapters provide aspects of the theory, technology, and applications of DNA-based markers. This book is intended to provide a solid base upon which students and researchers can build their understanding of genetic linkage and the applications of that knowledge. It also provides a statistical appreciation of DNA-based markers for the genetic dissection of qualitative and quantitative traits. Some of the newest ideas and concepts are expressed here for the first time. The locations of QTLs (Quantitative Trait Loci) for important agronomic and quality traits, as well as those that differentiate species, represent extremely useful knowledge. Genomic markers for regions with major effects for such traits as disease resistance already are being used for gene cloning and breeding purposes.

We recognize that genetic maps are a dynamic entity and will continue to expand and evolve over the years. At the same time, genetic map-based information is rigorous by its nature, and provides a set of basic tenets that should not change with future advancements. Thus, an understanding of the numerous basic tenets of mapping carefully described in these pages will establish a framework for future enhancement of the theory and applications of

this important and useful technology.

Clearly, our ability to genetically manipulate more than just a few plant species has increased several orders of magnitude with the advent of DNA-based markers. The future will be exciting.

We gratefully acknowledge the diligent efforts of each of the contributors to this volume for providing a clear statement of a difficult and still evolving subject. They represent pioneers in the field. Their interest and cooperation in making this volume a reality reflects a sincere and shared belief in the broad application of DNA-based markers in plants.

Ronald L. Phillips
Indra K. Vasil

1. Some concepts and new methods for molecular mapping in plants

BENJAMIN BURR

Biology Department, Brookhaven National Laboratory, Upton, NY 11973, U S A

Contents

1	Introduction	1	4	Syteny	5
2	The nature of naturally occurring polymorphism	2	5	Acknowledgement	6
3	PCR-based mapping methods	4	6	References	6

1. Introduction

The use of molecular markers is based on naturally occurring polymorphism whose magnitude has only recently been appreciated. Understanding the nature of naturally occurring polymorphism is a basis for designing strategies to exploit it for applied genetic practices. The purpose of this chapter is to express a personal view on a few key concepts and to draw attention to new technology.

One of the unheralded revolutions in biological thinking in the 1960's was the realization of the enormous amount of polymorphism present in natural populations. Although population geneticists should have realized this sooner from the variety of genetically inherited blood types that were known, most believed that there was little variation among wild-type alleles. The finding of Hubby and Lewontin (1966) that at least 30% of the loci they could assay in wild *Drosophila pseudoobscura* populations encoded electrophoretically polymorphic proteins changed forever the view that there was little polymorphism in natural populations, and seriously brought into question the notion that much of the change in allele frequencies was the result of natural selection. Consequently, the idea that there was widespread polymorphism in natural populations was soon confirmed for many species and provided geneticists with new tools for genetic analysis. Previously, our ability to map genes depended on synthesizing stocks multiply marked with mutations, which frequently reduced their viability. Now, we can make use of differences in wild-type alleles that potentially distinguish any two strains for linkage analysis.

2. The nature of naturally occurring polymorphism

It is important to understand something of the nature of naturally occurring genetic variation because it bears on the detection methods we employ. One of the most extensive studies was conducted by Kreitman (1983) who sequenced eleven wild-type *Alcohol dehydrogenase (Adh)* alleles of *Drosophila melanogaster*. He detected 43 basepair changes and ten length polymorphisms, all less than 40 basepairs. Only one of these changes led to a change in a charged amino acid in the encoded protein. One of the surprising findings from this study was that polymorphisms in introns, untranslated, and flanking regions (1.9%) were slightly less frequent than in coding regions (2.5%). In maize, most spontaneous mutations are the result of some form of insertion or deletion. For example, Wessler and Varagona (1985) found that 12 out of 17 spontaneous *waxy* mutations were associated with detectable (greater than 50 basepairs) deletions or insertions. On the other hand, when wild-type alleles are compared, small changes predominate. Ralston, English and Dooner (1988) sequenced three alleles of maize *bronze1*. In a region of over 3200 basepairs they detected 115 (77%) basepair changes, 28 (19%) deletions/insertions less than 20 basepairs, and only 7 (4%) major deletions or insertions. As is the case with *Drosophila Adh*, the frequency of polymorphism (1.5%) within the coding region was the same as non-translated and flanking regions. Of the 21 differences that distinguished the coding sequence of the two wild-type alleles, nine led to amino acid substitutions, but only three involved charged amino acids. To summarize, the lessons from a variety of studies indicate that base pair changes are more frequent than large rearrangements, that heterogeneity is not restricted by coding regions, and, not surprisingly, polymorphisms at the DNA level are much more frequent than are charge changes in proteins.

The level of variation in a species affects our ability to perform linkage analysis. The degree of polymorphism can be assessed by sequencing or making restriction maps of random wild-type alleles. Often this information is generated in the characterization of mutant alleles at a gene locus. The polymorphism present in natural populations differs markedly in different species (Evola et al. 1986). Differences at the DNA level vary from 1 to 2 basepairs per 1000 in humans to more than 40 per 1000 in maize. Even among plant species, these differences are remarkable. Shattuck-Eidens et al. (1990) sequenced six to eight RFLP alleles at each of four loci in maize and found an average of 1.2% variation. The spectrum of types of changes was similar to that observed at the *bronze1* locus. By contrast six alleles of three RFLP loci in melon (*Cucumis melo*) revealed only single base substitutions in each of two alleles at one locus for a level of variation of 0.02%, or about 50 times less than that observed for maize.

A commonly held view is that maize is exceptionally polymorphic because of the transposable element activity associated with the species. In addition to creating rearrangements, transposable elements can also generate small sequence changes. Schwarz-Sommer et al. (1985) pointed out that when trans-

posable elements are excised from a locus, they leave behind a 'footprint', most commonly a variant of the short sequence duplicated upon insertion, as a mechanism for generating diversity. When elements were inserted in exons, the revertant can sometimes have altered activity. However, Aquadro (1992) reminds us that among three closely related species, *Drosophila melanogaster* has eight times more copies of transposable elements than do *D. simulans* or *D. pseudoobscura*, but that the level of diversity in *D. melanogaster* is one-third to one-fifth the level of polymorphism in the other two species.

According to Nei (1983), if mutations are neutral, heterozygosity is a function of the effective species population size and the mutation rate. In fact, these factors tend to overestimate heterozygosity so Nei concluded that selection must be limiting heterozygosity to some extent. Helentjaris et al. (1985) compared the level of polymorphism as judged by RFLP among a number of plant species and noted that species that were self-pollinating showed much less variation than those that used out-crossing. This observation is consistent with Nei's hypothesis because self-pollination would be expected to reduce the effective population size. Hannah (pers. comm.) has noted out-crossing plants, because they are more likely to be heterozygous, can create new alleles by recombination and thus elevate the level of variation beyond that created by the mutation rate.

There are some regions of the genome that are significantly more polymorphic than single copy sequences. Assaying these regions is a way of increasing the level of detectable polymorphism between individuals. Tandem repeats of 15 to 60 basepair long sequences (mini-satellites or variable number tandem repeats, VNTRs) were originally discovered adjacent to unique sequence genes and found to account for an exceptionally high rate of RFLP (Jeffreys 1987; Nakamura et al. 1987). In fact, VNTRs form the basis for current methods of DNA typing in humans. While it may be difficult to identify new VNTRs, much simpler sequence repeats of one to four basepairs (microsatellites) that also exhibit a high degree of polymorphism (Webber and May 1989), are much easier to discover. Polymorphism of tandem repeats is thought to be generated by unequal crossing over or by slippage during DNA polymerization. Whatever the mechanism, the use of short tandem repeats has recently been a minor revolution in human and mammalian genetics in that it has provided ample polymorphism where it had been difficult to detect with more conventional molecular markers (Dietrich et al. 1992; Serikawa et al. 1992; Weissenbach et al. 1992). The tandem repeats are present throughout the genome; although, in humans, VNTRs may be clustered in the vicinity of telomeres, and microsatellites appear to be sparse in terminal regions (Weissenbach et al. 1992). Thus an oligonucleotide probe for one of these repeats will identify clones from other regions of the genome containing the repeat. Unique sequences that flank the tandem repeats can be used as highly polymorphic probes or for making PCR primers.

3. PCR-based mapping methods

Microsatellites are especially attractive for a number of reasons in addition to the fact that they are frequently highly polymorphic. First of all, the assay is PCR based. It is sufficient to merely separate the amplification products by electrophoresis to observe the results. This considerably reduces the time required to obtain a result compared with methods that are based upon Southern blotting. Secondly, the use of radioisotopes can be avoided because the size polymorphism between alleles is frequently large enough to be seen in agarose gels. However, it is possible to think of radionucleotide-based assays for microsatellites that could be automated and completely avoid the gel electrophoresis step. Methods that are based upon DNA sequence have another advantage over probe based strategies. The maintenance and distribution of probes has proven to be time consuming and frequently error prone. The only step in dissemination of sequence-based methods is to publish the DNA sequence along with the mapping results so that individual laboratories can synthesize their own oligonucleotides. Despite the high degree of polymorphism, microsatellite alleles seem to be surprisingly stable. Lander and colleagues (Dietrich et al. 1992) estimate a mutation rate of 4.5×10^{-5} in mice. In our own preliminary work we have observed that allelic differences are stable in recombinant inbreds that have been selfed eight times beyond the F_2 . The major disadvantage of microsatellites is the cost of establishing polymorphic primer sites and the investment in synthesizing the oligonucleotides. Once these costs are incurred, however, the long term use of this method will pay off in reduced man power and materials costs.

While microsatellites are well established for human and mammalian genetics, there are also two recent surveys of their practicality in plants (Akkaya et al. 1992; Morgante and Olivieri 1993). Both of these studies demonstrated the segregation of microsatellites as co-dominant markers and showed that this type of polymorphism is prevalent in soybean, a species in which it has been difficult to find polymorphic markers in the past. There are well established methods of finding microsatellites in phage libraries by screening with oligonucleotide probes. But a quicker, if limited, approach is to examine sequence databanks for their presence. The major result, noted by Morgante and Olivieri (1993), is that while AC repeats are the most frequent class in humans and rats (Beckmann and Weber 1992), it is the AT repeat that is the most prevalent in higher plants.

It is interesting to observe that while the use of tandem repeats dominates animal work, another method based on random priming is used almost exclusively in plant molecular mapping. This technique was developed in two laboratories (Williams et al. 1990; Welch and McClelland 1990) and is frequently referred to with the acronym RAPD. It depends on the observation that single short oligonucleotide primers can frequently recognize similar sequences that are opposed to each other at distances close enough for the intervening sequence to be amplified in the PCR. Entire maps have been made using these primers exclusively (Reiter et al. 1992; Chaparro et al. 1992). One of

the most intense uses of RAPDs is for finding new markers that are tightly linked with a specific locus. Because of the availability of random oligonucleotides and the relatively easy assay to look for linkage, it is a simple matter to screen many loci rather rapidly. The application of an ingenious idea makes this process even simpler. Bulk segregant analysis (Michelmore et al. 1991) uses two groups of individuals, optimally each group is homozygous for alternate alleles that govern the trait. If members of a segregating population are pooled on the basis of their phenotype, then the groups can be expected to be heterozygous at all unlinked loci but markedly skewed in the direction of one or the other parental allele for a linked locus. Giovannoni et al. (1991) demonstrated the use of this technique to target a region after individuals had been pooled on the basis of their RFLP phenotype. As powerful as it is, there are two properties of RAPD analysis that limit its application. Priming polymorphism appears to be based on mismatches with target sequences so that alleles are either present or absent. Since there is no guarantee that the dominant allele will be present in a second population, it is not always possible to use a mapped RAPD locus with new parents or to map a RAPD, that has been found to be linked with an interesting locus, in a new population. One way around this problem is to convert RAPD bands to RFLPs by cloning the amplification product. This has worked in species with small genomes like tomato and rice, but may prove very difficult in plants with larger genomes and more repeated DNA because there is no preference for amplifying single copy sequences. An alternative approach is to sequence the ends of a RAPD band so that stable primers can be synthesized to amplify this band preferentially in several genotypes (Paran and Michelmore 1993).

Two final PCR-based methods deserve mention. Allele-specific PCR is based on the choice of primers so that one allele or set of alleles will not amplify and another allele, the mutant one is interested in following for instance, will. Shattuck-Eidens et al. (1991) have demonstrated the technique for a *wx* allele of maize and have shown that non-isotopic hybridization techniques are sufficient to distinguish the mutant from the wild-type allele in dot blots of the amplification reactions, thus avoiding a gel electrophoresis step. It has been known for some time that the conformation of single stranded DNA during gel electrophoresis is dependent on DNA sequence and can be altered by single nucleotide substitutions to produce altered mobility. Single strand conformation polymorphism can be demonstrated for short (200 to 400 nucleotide) PCR products to demonstrate polymorphism (if it exists) when other methods have failed (Orita et al. 1989).

4. Synteny

A recent and important observation in the use of plant molecular markers is the finding that many distantly related species have co-linear maps for portions of their genomes. This was first demonstrated for members of the Solanaceae

(Bonierbale et al. 1988; Tanksley et al. 1988) and has been extended to a demonstration of extensive co-linearity between maize and sorghum (Hulbert et al. 1990; Whitkus et al. 1992) and more remarkably between maize and rice (Ahn 1993). There are indications that these relations might be conserved in even more distantly related plants because the duplicated sucrose synthase loci are on the same linkage group in both maize and wheat (Marana et al. 1988). The obvious result of this work is that, once careful interspecific comparisons have been made, predictions for gene locations mapped in one species can be made in the second. Since both sorghum and rice have smaller genomes than maize, it may be feasible to use one of these species to clone a gene sought in maize. In the short run, regions that are sparsely mapped in one species might be filled by markers from the molecular map of the second species.

5. Acknowledgement

Work was performed at Brookhaven National Laboratory under the auspices of the U.S. Department of Energy Office of Basic Energy Sciences.

6. References

- Aquadro, C.F. (1992) Why is the genome variable? Insights from *Drosophila*. *Trends Genet.* 8: 355–362.
- Ahn, S.N. (1993) Comparative mapping of the rice and maize genomes and mapping quality genes of rice using restriction fragment length polymorphism (RFLP) markers. Ph.D. thesis, Cornell University.
- Akkaya, M.S., Bhagwat, A.A. and Cregan, P.B. (1992) Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics* 32: 1131–1139.
- Beckmann, J.S. and Weber, J.L. (1992) Survey of human and rat microsatellites. *Genomics* 12: 627–631.
- Bonierbale, M.W., Plaisted, R.L. and Tanksley, S.D. (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* 120: 1095–1103.
- Chaparro, J., Wilcox, P., Grattapaglia, D., O'Malley, D., McCord, S., Sederoff, R., McIntyre, L.M. and Whetten, R. (1992) Genetic mapping of pine using RAPD markers. In: *Advances in Gene Technology: Feeding the World in the 21st Century*. Miami Winter Symposium, Miami, FL.
- Dietrich, W., Katz, H., Lincoln, S.E., Shin, H.-S., Friedman, J., Dracopoli, N.C. and Lander, E.S. (1992) A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* 131: 423–447.
- Evola, S.V., Burr, F.A. and Burr, B. (1986) The suitability of restriction fragment length polymorphisms as genetic markers in maize. *Theor. Appl. Genet.* 71: 765–771.
- Giovannoni, J.J., Wing, R.A., Ganai, M.W. and Tanksley, S.D. (1991) Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucl. Acids Res.* 19: 6553–6558.
- Helentjaris, T., King, G., Slocum, M., Siederstrang, C. and Wegman, S. (1985) Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. *Plant Mol. Biol.* 5: 109–118.
- Hubby, J.L. and Lewontin, R.C. (1966) A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* 54: 577–594.

- Hulbert, S H , Richter, T E , Axtell, J D and Bennetzen, J L (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes Proc Natl Acad Sci U S A 87 4251-4255
- Jeffreys, A J (1987) Highly variable minisatellites and DNA fingerprints Biochem Soc Trans 15 309-317
- Kreitman, M (1983) Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster* Nature 304 412-417
- Marana, C , Garcia-Olmedo, F and Carbonero, P (1988) Linked sucrose synthase genes in group-7 chromosomes in hexaploid wheat (*Triticum aestivum* L) Gene 63 253-260
- Michelmore, R W , Paran, I and Kesseli, R V (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis a rapid method to detect markers in specific genomic regions using segregating populations Proc Natl Acad Sci U S A 88 9828-9832
- Morgante, M and Olivieri, A M (1993) PCR-amplified microsatellites as markers in plant genetics Plant J 3 175-182
- Nakamura, Y , Leppert, M , O'Connell, P , Wolf, R , Holm, T , Culver, M , Martin, C , Fugimoto, E , Hoff, M , Kuhlman, E and White, R (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping Science 235 1616-1622
- Nei, M (1983) Genetic polymorphism and the role of mutation in evolution In M Nei and R K Koehn (eds), Evolution of Genes and Proteins, pp 165-190 Sinauer Assoc , Sunderland, MA
- Orta, M , Susuki, Y , Sekiya, T and Hayashi, K (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction Genomics 5 874-879
- Paran, I and Michelmore, R W (1993) Sequence characterized amplified regions (SCARs) as a codominant genetic marker in *Lactuca* spp Theor Appl Genet 85 985-993
- Ralston, E J , English, J J and Dooner, H K (1988) Sequence of three *bronze* alleles of maize and correlation with the genetic fine structure Genetics 119 185-197
- Reiter, R S , Williams, J , Feldman, K A , Rafalski, J A , Tingey, S V and Skolnick, P A (1992) Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplification polymorphic DNAs Proc Natl Acad Sci U S A 89 1477-1481
- Schwarz-Sommer, Z , Gierl, A , Cuyper, H , Peterson, P A and Saedler, H (1985) Plant transposable elements generate the sequence diversity needed in evolution EMBO J 4 591-597
- Serikawa, T , Kuramoto, T , Hilbert, P , Mori, M , Yamada, J , Dubay, C J , Lindpainter, K , Ganten, D , Guenet, J -L , Lathrop, G M and Beckmann, J S (1992) Rat gene mapping using PCR-analyzed microsatellites Genetics 131 701-721
- Shattuck-Eidens, D M , Bell, R N , Neuhausen, S L and Helentjaris, T (1990) DNA sequence variation within maize and melon observations from polymerase chain reaction amplification and direct sequencing Genetics 126 207-217
- Shattuck-Eidens, D M , Bell, R N , Mitchell, J T and McWorter, V C (1991) Rapid detection of maize DNA sequence variation GATA 8 240-245
- Tanksley, S D , Bernatzky, R , Lapitan, N L and Prince, J P (1988) Conservation of gene repertoire but not gene order in pepper and tomato Proc Natl Acad Sci U S A 85 6419-6423
- Weber, L and May, P E (1988) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction Am J Hum Genet 44 388-396
- Weissenbach, J , Gyapay, G , Dib, C , Vignal, A , Morissette, J , Millasseau, P , Vaysseix, G and Lathrop, M (1992) A second-generation linkage map of the human genome Nature 359 794-801
- Welsh, J and McClelland, M (1990) Fingerprinting genomes using PCR with arbitrary primers Nucl Acids Res 18 7213-7218
- Wessler, S R and Varagona, M J (1985) Molecular basis of mutations at the waxy locus of maize Correlation with the fine structure genetic map Proc Natl Acad Sci U S A 82 4177-4181
- Whitkus, R , Doebley, J and Lee, M (1992) Comparative genome mapping of sorghum and maize Genetics 132 1119-1130
- Williams, J G K , Kubelik, A R , Livak, K J , Rafalski, J A and Tingey, S V (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers Nucl Acids Res 18 6531-6535

2. RFLP technology

GARY KOCHERT

Department of Botany, University of Georgia, Athens, GA 30602, U S A

Contents

1	Introduction	8	3 1 2	Minisatellites	21
2	Conventional RFLP analysis	10	3 1 3	Microsatellites	23
2 1	DNA isolation	11	3 1 4	Mapping with VNTR loci	23
2 2	Restriction enzyme digestion and agarose electrophoresis	12	3 2	Single strand conformation polymorphisms	25
2 3	RFLP probe libraries	13	3 3	Single primer PCR methods	26
2 4	Probe labelling	15	3 4	Locating markers in defined chromosomal segments	28
2 5	Mapping populations	16	4	Future directions	30
3	Recent technological advances	19	5	References	33
3 1	PCR based markers	20			
3 1 1	VNTR loci	21			

1. Introduction

In the early part of the 20th Century, scientists discovered that Mendelian 'factors' controlling inheritance, which we now call genes, were organized in linear order on cytogenetically defined structures called chromosomes. Shortly thereafter the first chromosome map was produced by Sturtevant with segregation data derived from studies on *Drosophila* (Crow and Dove 1988). The markers on this first genetic map were phenotypic traits scored by visual observation of morphological characteristics of the flies.

There has been continuous interest in chromosome markers since the first genetic map was produced. Chromosome maps of various types have been produced for a diverse array of organisms, including economically important plants or animals or those organisms which have been useful as model systems. Until recently, the markers on these maps were either genes or morphological features of the chromosomes themselves as determined by cytogenetic studies. The linear order of genes on chromosomes can be determined by phenotypic or physiological scoring of segregating progeny from a sexual cross. On the resultant map, distances between genes are given as 'genetic distance' reflecting

the probability of a crossover event occurring between the markers in question. Such distances can only be determined by scoring the phenotype produced by gene action of progeny from a segregating population. Cytological maps can be produced by direct observation of the chromosomes, and distances between markers are determined in terms of physical units, such as micrometers.

The construction of chromosome maps has, until recently, proceeded slowly for several reasons. One of the most important of these is the lack of polymorphic markers for both genetic and physical maps. To construct a genetic map based on conventional genes it is necessary to have two forms (alleles) of the same gene locus. In classical genetics these are often referred to as 'wild type' and 'mutant' forms. Mutants could be searched for in natural populations or could be produced by mutagenesis. Often it was necessary to make crosses, allelism tests, and perform segregation analysis to demonstrate whether the different phenotypes observed were really due to different alleles at the same locus or different loci. After different alleles of two or more loci have been found, it is necessary to assemble these into two parents suitable for a mapping cross. After crossing the parents to produce a suitable mapping population such as a backcross or F_2 , the genes can be mapped by segregation analysis. However, the only way to assemble genes into suitable mapping parents is by a series of genetic crosses to construct marker stocks. These require a great deal of time and effort resulting in slow progress. It is difficult to assemble a large number of loci into any one marker stock, therefore only a few genes can be mapped in any one cross. Thus the construction of genetic maps is a long-term effort requiring much labor and many crosses. For these reasons, detailed genetic maps showing the location of genes on chromosomes have been produced for only a few organisms.

A diploid organism contains one set of chromosomes from each of its parents. When that organism undergoes meiosis, these chromosomes undergo recombination so that each chromosome in the resultant gametes contains alternating chromosome segments derived from each of the parent chromosomes (Fig. 1). Genetic maps are constructed by determining the probability of two chromosome segments from the same parents becoming separated by recombination. In conventional mapping the only way to determine whether two chromosome segments have cosegregated or have become separated is by indirect observation. Since the chromosome segments can not be directly scored it is necessary to infer their behavior by scoring for the presence or absence of alleles known to be on the respective segments. This is a cumbersome way to score for chromosome segments because the alleles themselves can not be observed. It is necessary to depend on gene expression and scoring of the resultant phenotype to infer the presence of a certain allele and hence a certain chromosome segment.

A major breakthrough occurred when it was realized that genetic maps could be constructed by using pieces of chromosomal DNA as direct markers for the segregation pattern of chromosome segments (Botstein et al. 1980). The first such DNA markers to be utilized were fragments produced by restriction

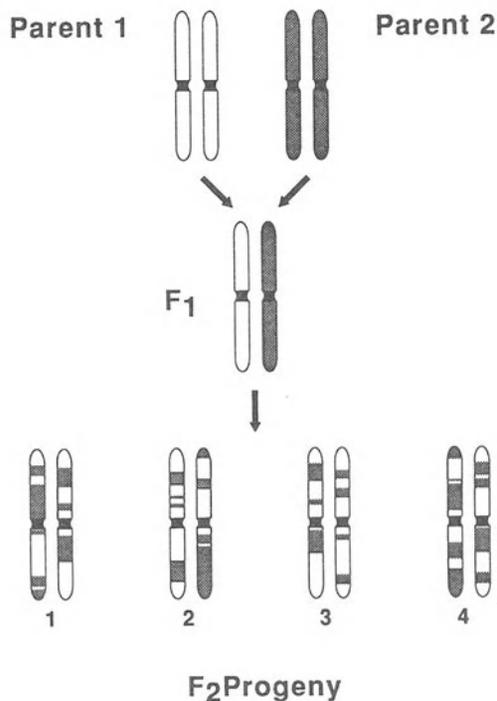


Fig. 1. Distribution of chromosome segments by meiosis in a cross of inbred, homozygous parents. Only one pair of homologous chromosomes is shown.

enzyme digestion. It was soon found that restriction fragments from a given chromosomal locus often varied in size (or length) in different individuals of the same or different species, and these differences were designated restriction fragment length polymorphisms or RFLPs. RFLPs have their origin in base sequence changes or in DNA rearrangements, and are naturally occurring, simply inherited, Mendelian characters. Thus they can be used to construct genetic maps in which the markers are RFLPs.

RFLPs have proven to be abundant in most organisms, and a virtually unlimited number can be mapped in any one cross. Thus it is not necessary to laboriously construct marker stocks for genetic mapping. Since RFLPs are stable attributes of the DNA itself, and scoring does not depend on gene expression, DNA for mapping can be isolated from any part of the plant at any stage of the life cycle.

2. Conventional RFLP analysis

In its original form, RFLP analysis consisted of DNA isolation from a suitable set of plants, digestion of the DNA with a restriction enzyme, separation of the

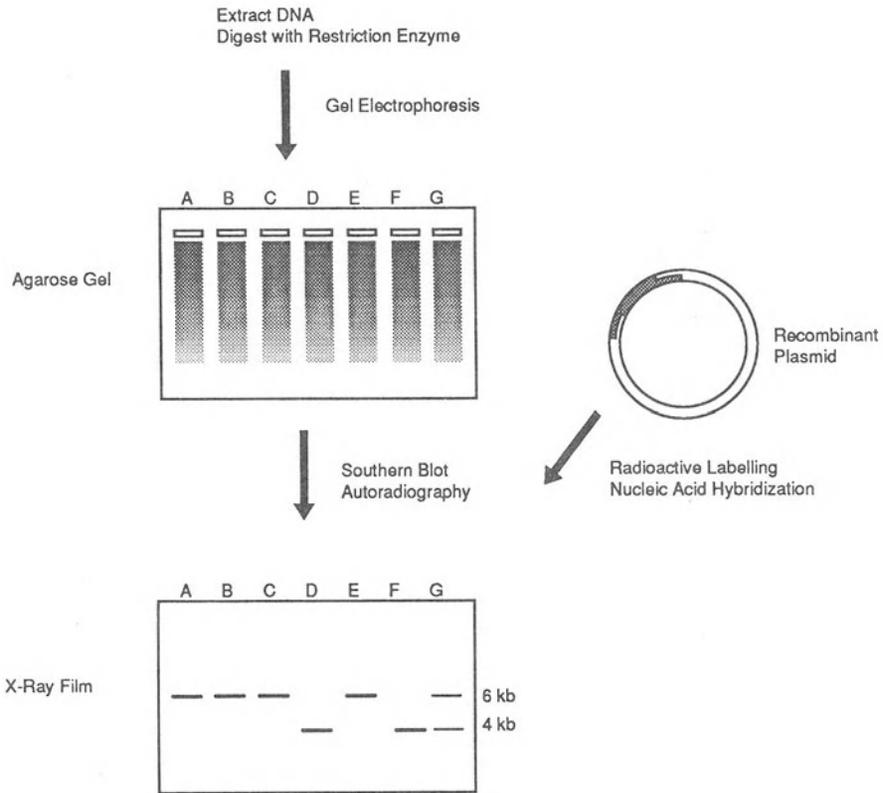


Fig 2 Conventional RFLP analysis

restriction fragments by agarose gel electrophoresis, transfer of the separated restriction fragments to a filter by Southern blotting, detection of individual restriction fragments by nucleic acid hybridization with a radioactively labelled cloned probe, and scoring of RFLPs by direct observation of autoradiograms (Fig. 2). Changes have taken place in several aspects of this process. For example the development of polymerase chain reaction (PCR) techniques has made possible new avenues of investigation. The salient details of conventional RFLP analysis will first be described, and this will be followed by a discussion of newer approaches to mapping. Earlier reviews of this general topic have been provided by Helentjaris et al. (1985), Landry and Michelmore (1987), and Tanksley et al. (1988b).

2.1. DNA isolation

For routine analysis of RFLPs it is only necessary to purify plant DNA enough so that it will digest reproducibly with restriction enzymes, and the resultant preparations can be satisfactorily separated by gel electrophoresis. In plants, the

greatest problems are likely to be caused by the use of tissues that contain an abundance of carbohydrates, glycoproteins, or secondary products which tend to co-purify with the plant DNA and prevent proper digestion and electrophoresis. With most grasses, such as rice, maize, and bamboo, very simple isolation procedures can be used (McCouch et al. 1988; Friar and Kochert 1991). The most common starting material is young leaves. These can be either fresh, lyophilized, dried in an oven, or in some cases dried at room temperature. If the leaves are fresh, they are ground into a powder with a mortar and pestle in liquid nitrogen; if lyophilized or dried, a mechanical grinding mill can be used. The powder is then suspended in an extraction buffer. Many sorts of extraction buffers have been used, but most contain some or all of the following components:

1. A buffer to maintain the pH at around 8.0.
2. A salt, such as NaCl to aid in dissociating proteins from the DNA.
3. Detergents such as SDS or sarkosyl to solubilize plant membranes.
4. Some means of rapidly inactivating DNase. Since DNase requires magnesium ions for activity, EDTA is often added to sequester the magnesium ions. Some protein denaturant, such as phenol, chloroform, or urea may also be used. Any added detergents will also aid in inactivation of DNase, as will incubation of the extract at elevated temperatures.

Most proteins are then removed by treatment with chloroform, phenol, or a protease, and the DNA is reprecipitated with alcohol and solubilized in buffer. Ribonuclease treatment can be used at this point to remove RNA, but RNA does not interfere with restriction enzyme action or electrophoresis. Such total DNA preparations contain a mixture of nuclear DNA and organellar DNA, and they can be used for analysis of either nuclear or chloroplast RFLPs.

With other plants, including legumes such as peanuts or alfalfa, special steps must be taken to purify the DNA from carbohydrates or glycoproteins. This can be accomplished by breaking open the cells of fresh tissue in a blender or Polytron homogenizer and making a crude nuclear preparation for DNA isolation (Murray and Thompson 1980; Kochert et al. 1991). The supernatant from the nuclear preparation contains the bulk of the carbohydrates and glycoproteins. Selective precipitation by cetyl-trimethyl-ammonium bromide (CTAB) of nucleic acids from solutions containing polysaccharides (Murray and Thompson 1980; Saghai-Marooof et al. 1984; Rogers and Bendich 1988) can also be used. In only the most recalcitrant cases, or those cases where an accurate estimate of DNA quantity is needed, it is necessary to purify the DNA by more complex methods such as equilibrium centrifugation in cesium chloride gradients.

2.2. *Restriction enzyme digestion and agarose electrophoresis*

Enzyme digests are conveniently performed in volumes small enough (about 30 μ l) so that the digest can be loaded directly into the well of an agarose gel. The

choice of restriction enzyme to be used for detection of RFLPs is usually made on the basis of the cost of the enzyme and its efficacy for detecting variability. RFLP variability in plants can be caused by (1) base sequence changes which add or eliminate restriction sites, (2) rearrangements such as insertions or deletions, or (3) unequal crossing over or replication slippage (Schlötterer and Tautz 1992) which creates variation in the number of tandem DNA sequence repeats at a minisatellite or microsatellite locus (see below). In those cases where a study has been made, it appears that most RFLP variability in plants is caused by genome rearrangements rather than nucleotide sequence change (Landry et al. 1987; Apuya et al. 1988; McCouch et al. 1988; Miller and Tanksley 1990). Evidence for this view is derived from (1) observations that '6-cutter' restriction enzymes (i. e. those restriction enzymes which recognize a 6 base pair site in the genome) reveal more RFLPs than do '4-cutters' in spite of the fact that, on average, 4-cutters survey more DNA sequence than do 6-cutters (for example, in a 16 kb stretch of DNA, a 4-cutter enzyme would be expected to recognize about 64 sites or 256 nucleotide pairs while a 6-cutter would detect 5 sites or 30 nucleotide pairs). (2) the amount of RFLP variation detected correlates with the average length of the restriction fragments produced by that enzyme, and (3) RFLPs detected by one enzyme tend to be detected by multiple enzymes.

Following digestion with a restriction enzyme, DNA samples are separated by agarose electrophoresis. The amount of DNA required per lane of the gel will vary from 2 or 3 μg for plants with a small genome (*Arabidopsis*, rice) to as much as 15 μg for those with large genomes (maize, wheat). Genome sizes for many plants have been tabulated in the papers of Bennett and Smith (1976, 1991), Bennett et al. (1982) and Arumuganathan and Earle (1991).

The DNA is then transferred to nylon membranes by Southern (1975) transfer. There are differences of opinion about how nylon membranes should be handled after DNA transfer to ensure maximum use. Some researchers simply air dry them, others bake them at 80 °C or 'crosslink' the DNA to the membrane by UV treatment, and there is no clear consensus. However, under some conditions membranes can be used for analysis of nuclear RFLPs 30 or more times, and this is an important consideration when calculating the time and expense involved in an RFLP study. Following use for nuclear RFLP analysis, the membranes may still be useful for analysis of sequences which are present many times in a cell, such as rRNA genes, nuclear repeated sequences or chloroplast sequences.

2.3. RFLP probe libraries

Both cDNA and random genomic libraries have been used as sources of probes for RFLP mapping, and each type of library has advantages and disadvantages. Random genomic libraries are easily constructed by digesting total DNA extractions with a restriction enzyme and cloning into a suitable vector, usually a plasmid. Because of the interspersed nature of many repeated DNA sequences in plants, small mapping clones (0.5–2 kb) are utilized because most larger

clones will contain repeated DNA. Every type of sequence will be included in random genomic libraries, but single or low-copy number clones are most useful for routine RFLP map construction. Interspersed repeated sequences which are present in high copy number will detect a large number of restriction fragments on autoradiograms from Southern blots and will often appear as a smear. Interspersed repeats of lower copy number will produce autoradiograms with resolvable bands which can be mapped. However, loci mapped with such probes may be of limited utility for an RFLP map which will be used with different populations of the same species. For example, if a probe detects 12 restriction fragments in a given segregating population and two are polymorphic, these can be mapped. However, if the map is transferred to another segregating population, a different subset of the fragments could be polymorphic, and the chromosomal location of these would be unknown. For these reasons it is best to construct maps with single copy sequences, unless the goal of the project is to study genome evolution.

Repeated sequences can be excluded from genomic libraries or the libraries can be screened after construction to select low copy number sequences. Repeated sequences can be partially excluded from genomic libraries if the libraries are constructed using *Pst*I digestions (Figdore et al. 1988; Liu et al. 1990; Miller and Tanksley 1990). Repeated sequences in plants are frequently methylated at C sequences, and *Pst*I is methylation-sensitive and will not cleave the DNA if the C at the 5' end of its recognition site (5'-CTGCAG-3') is methylated. Thus many repeated sequences are not cleaved and remain as large fragments in *Pst*I restriction digests. These large fragments ligate into plasmids and transform into bacteria with low efficiency and thus are effectively excluded from the resultant library, which is enriched for low copy number sequences. Screening of libraries to select low copy number clones can be carried out using ³²P labelled total DNA as a hybridization probe on filters prepared by colony lifts or from minipreps of plasmids. Clones which reveal a hybridization signal under these conditions are repeated sequence clones and are not tested further. For unknown reasons, however, some repeated sequence clones will escape this screening and will only be revealed in subsequent steps (Landry et al. 1987). Those fortunate enough to be working in plants with small amounts of repeated DNA, such as *Arabidopsis* or mung bean, often dispense with library screening and simply discard the infrequent repeated sequences when they are encountered later in the screening process.

cDNA libraries are more difficult to construct than random genomic libraries, but seem to be gaining in popularity as sources of RFLP probes. Part of this is due to the knowledge that one is mapping actual genes when a cDNA library is employed. With the rapid growth in the number of published sequences of known genes and the recent trend toward patenting new cDNAs in the hope that they will eventually be useful (and profitable), there is hope that any mapped cDNA can be eventually identified by sequence comparisons. Libraries made from cDNAs contain fewer repeated sequences so that preliminary screening is not usually necessary.

In any given plant system it is not clear whether cDNA libraries or random genomic libraries will reveal more polymorphism for RFLPs. As a general principle it might be considered that spaces between genes, which would be the predominant clones encountered in a random genomic library might be more variable than the genes themselves, which should be under more selection. However, this has not always proven to be the case, but only a few studies have been done. In lettuce, Landry et al. (1987) found that cDNA clones detected 2.5 times more polymorphism than did random genomic clones. In tomato, random genomic libraries were constructed from *Pst*I and *Eco*RI digests and compared to a cDNA library. Probes from the cDNA library detected 25% more unique restriction fragment patterns than did those from the *Pst*I library and almost 50% more than did probes from the *Eco*RI library (Miller and Tanksley 1990). cDNA probes used as hybridization probes do not only detect variation in the coding region of the corresponding gene. RFLP polymorphisms in regions flanking the gene and in introns of the gene are also detected. Both these areas would be expected to be more polymorphic than the coding region of the gene itself, and they are probably largely responsible for the polymorphisms detected when cDNAs are used as RFLP probes.

No underlying genome organization of coding sequences versus non-coding regions has been revealed by mapping low copy number random genomic clones and cDNA clones in the same organism. A random-appearing interspersion of sequences homologous to the two types of clones appears to be the rule in maize (Helentjaris 1987), tomato (S. Tanksley, pers. comm.) and rice (S. Tanksley, unpub.; see also Chapters 17 and 20, this volume).

2.4. Probe labelling

Most RFLP work has been done using probes radioactively labelled with ^{32}P , because the short half-life of ^{32}P allows probes to be labelled to high specific activity. This, coupled with the high energy of the beta particle emitted by ^{32}P , results in reasonably short autoradiogram exposure times for the detection of single-copy nuclear sequences. Enhancement of the signal can also be achieved by using enhancing screens. Where high resolution of bands on autoradiograms is necessary, such as DNA sequencing or DNA fingerprinting (Johnson and Th'ng 1991), ^{35}S can be used. Longer exposures may be needed, but the lower energy beta emitted will produce bands which are better resolved. Labelling of probes for RFLP analysis is usually carried out with only one labelled nucleotide in the reaction mix, although if higher specific activities are required, more than one labelled nucleotide can be used.

To achieve labelling *in vitro* with an isotope, nick-translation (Rigby et al. 1977) or random primer labelling (Feinberg and Vogelstein 1984) is used, and they achieve higher specific activity than end-labelling. No net gain of probe sequences is achieved with nick-translation, so a relatively large amount of starting clone is needed. The probe must be double-stranded to serve as a substrate for nick-translation. Random primer labelling requires a single-

stranded substrate, so probes are heated to denature the DNA and then quick-cooled. Net synthesis of labelled probe occurs with random primer labelling, so smaller amounts of starting material are needed, and very high-specific activities can be achieved.

For some purposes it is not necessary to isolate the cloned DNA insert from the vector, and nick-translation or random primer labelling of an entire plasmid containing the insert has been used. However, it is usually better to isolate the insert and label it. This was formerly somewhat tedious and depended on isolating the insert from pieces cut from agarose gels. Insert isolation is now easily achieved by PCR using universal primers produced from vector sequences flanking the insert. Net synthesis of sequence occurs, so the insert can be produced from small amounts of starting material such as lambda bacteriophage plaques.

Non-radioactively labelled probes can also be used in RFLP analysis, and they are desirable for reasons of safety and for use in areas of the world where isotopes are difficult to obtain or use. Labelling of the probe with a modified nucleotide, such as one containing digoxigenin or biotin, is carried out by nick translation or random primer labelling. The labelled probe is usually stable for an extended period and can be used for several experiments. This is a major advantage over radioactively labelled probes. The labelled probe is then hybridized to filters prepared from Southern blots in the same fashion as are radioactively labelled probes. Detection of the modified nucleotide in the hybridized bands then depends on some sort of enzyme-coupled immunological binding process, or in the case of biotin, binding with avidin, leading to the production of a colored product. One system uses chemiluminescence which can be produced if an enzyme like horseradish peroxidase is covalently coupled to the probe and then detected by adding luminol and peroxide.

It is not completely clear why non-radioactively labelled probes have not been more widely used for RFLP analysis. Early protocols suffered from a lack of sensitivity and reproducibility, and those which produced a colored precipitate to mark the site of hybridization rendered the filters unusable for subsequent hybridizations. However, these limitations are being overcome, and in some plants results comparable to those obtained with radioactive probes can be achieved (Ishii et al. 1990). In developing countries, the use of non-radioactive probe technology may become more widespread, if hybridization technology is not eclipsed by PCR-based methods (see below).

2.5. Mapping populations

The choice of a suitable mapping population for the construction of molecular maps in plants will depend on the breeding system of the particular plant. So far most RFLP maps have been produced using F_2 populations or backcross populations derived from crosses between inbred parent lines. Even the maps of maize, which is normally an outcrossing plant, were produced from inbred parents. Using inbred parents simplifies genetic analysis because the phase

(coupling or repulsion) of the markers is completely known. F_2 populations provide more mapping information for a given number of plants when codominant markers are analyzed, since two recombinant chromosomes can be scored in each plant (Allard 1956; Tanksley et al. 1988b; Reiter et al. 1992). F_2 populations provide a sex-averaged map because chromosomes from both the male and female parent are scored. Backcross populations can provide a male or female map depending on which sex was the recurrent parent. Male and female genetic maps generally vary in length as witnessed by studies in humans (Donis-Keller et al. 1987) and plants (De Vicente and Tanksley 1991).

One of the greatest advantages of molecular markers is that a virtually unlimited number of markers can be mapped using a single segregating population. As long as the same set of F_2 or backcross plants is used, the database of mapped markers accumulates. When a map containing 100–200 well-dispersed markers has been constructed, virtually any new marker will be linked to one that has been previously mapped. Thus the database of mapped markers for a given segregating population becomes a valuable resource for mapping of new markers. A permanent mapping population and its accumulated mapping data can be transferred to a new research group, and the new group can begin mapping their own molecular or conventional markers relative to the framework map which has already been produced. If the mapping population is lost, the mapping information must be transferred to a new population by scoring some of the same markers that have been previously mapped. This wastes time and resources, so there has been considerable interest in the construction of permanent mapping populations.

Backcross and F_2 populations constructed from inbred lines are segregating populations and are not a permanent resource in most plants. However, perennial plants or plants which can be reproduced asexually, such as alfalfa (Brummer et al. 1991) or a rice population derived from an interspecific cross where one of the parents is perennial (M. Causse, unpub.), constitute a permanent mapping population. It is possible to reconstitute F_2 mapping populations by selfing each F_2 plant, growing F_3 plants from the resultant seed, and combining tissue from several of the F_3 plants for DNA extraction. The bulk DNA samples thus produced from each F_2 can be used in mapping just as was the original F_2 DNA. This sort of bulking of F_3 plants must be routinely done for species such as *Arabidopsis thaliana*, where the individual F_2 plants are too small to yield enough DNA for conventional RFLP analysis.

Although permanent mapping populations can be transferred by making cuttings in some plant species, the most useful type of permanent mapping population is one that is homozygous and can be reproduced by seeds. Seeds are easily transported, and there is often less problem in transferring seeds than in transferring plant cuttings, particularly where international boundaries are involved. Mapping populations that 'breed true' can be constructed in some plants by single-seed descent with selfing of individual F_2 or backcross plants. Six or more generations of selfing are required to drive the plants toward homozygosity, the resultant recombinant inbred (RI) lines will have become

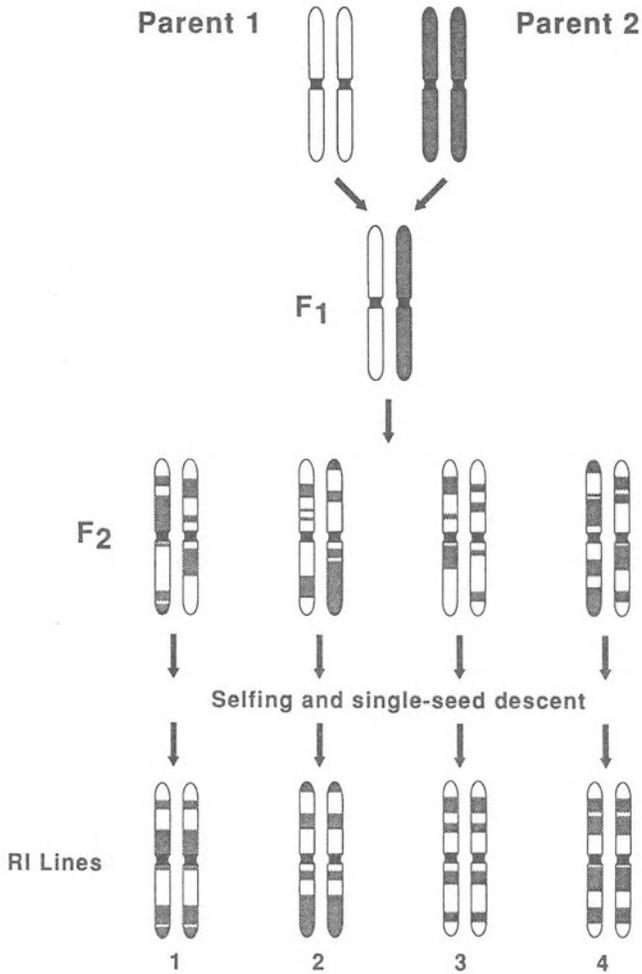


Fig. 3. Production of recombinant inbred lines by selfing and single-seed descent.

largely homozygous and can be propagated by seeds. Each chromosome of a recombinant inbred line contains alternating segments derived from each of the mapping parents (Fig. 3). Markers which are linked are more likely to stay together during the crossing and selfing which produces the lines and will show cosegregation. To generate mapping data, the lines are probed with a series of molecular markers, and each marker can be scored as a homozygote for one or the other of the parents in each of the lines (Burr et al. 1988; Burr and Burr 1991; Reiter et al. 1992). For markers which are inherited in a dominant-recessive fashion, such as the majority of RAPD (random amplified polymorphic DNA) markers, recombinant inbred lines provide as much efficiency in mapping as do segregating populations, and by using RAPD markers (see below), genetic maps can be constructed very quickly (Reiter et al. 1992). With F₂ mapping

populations, the results of meiosis in one generation are being scored, but recombinant inbred lines are the results of a series of meioses which give more opportunities for recombination. Thus a map derived from recombinant inbreds will have higher resolution than one derived from an F_2 or backcross population (Burr et al. 1988).

Homozygous mapping populations similar to recombinant inbreds can be produced in one step in those plants where doubled haploids can be produced by anther culture of an F_1 (Rivard et al. 1989; Powell et al. 1992). The resultant plants are homozygous and the chromosomes contain alternating segments derived from each of the parents just as do RI lines. However, only one meiosis has occurred so there is no increase in resolution over using an F_2 population, but the population will breed true and is permanent. Chromosome elimination such as that which occurs in some inter-specific or inter-generic crosses (Kasha and Sadasivaiah 1971) could also be used to produce homozygous populations by doubling the resultant haploids. For other plants it should be possible to use anther culture to produce haploid callus. This tissue can be typed in the same way as can recombinant inbred lines and maps could be rapidly constructed. However, changes in RFLP patterns have been reported to occur in tissue culture (Roth et al. 1989; Muller et al. 1990; Brown et al. 1991), and the degree to which these might affect the validity of maps constructed from doubled haploids or haploid callus has not been rigorously tested. In the case of barley, however, less than 0.5% nonparental RFLP patterns were found, indicating a low level of gametoclonal variation (Heun et al. 1991).

In cases where inbred lines are not available or cannot be constructed, such as in plants which exhibit a high degree of inbreeding depression, F_1 or backcross populations can be used (Bonierbale et al. 1988). For discussions of appropriate algorithms for use in a variety of mapping situations such as heterozygous populations or polyploids, see Ritter et al. (1990) and Wu et al. (1992).

3. Recent technological advances

Conventional RFLP analysis is limited by the relatively large amount of DNA required for restriction digestion, Southern blotting and hybridization plus the requirement for radioactive isotopes and autoradiography. These factors make conventional RFLP analysis relatively slow and expensive. The development of new methods to perform analysis with molecular markers has been the focus of many recent studies, and most of these are based on PCR amplification of genomic DNA and detection of differences in the PCR products as the basis for construction and utilization of molecular maps.

3.1. PCR based markers

Fragments of genomic DNA suitable as genetic markers can be produced by PCR amplification. This can be done by synthesizing PCR primers to uniquely amplify portions of the sequence of known genes (D'Ovidio et al. 1990) or mapped RFLP markers (Williams 1991). In the most favorable case, the PCR products from the two mapping parents will be of different size, will be inherited as codominant markers, and the resultant polymorphisms can be directly observed by running the products on an agarose gel. No Southern blots, DNA hybridization, or autoradiography are necessary. However, this direct approach will detect length polymorphisms less often than will conventional RFLP analysis because a smaller portion of the genome is surveyed. PCR products can be a maximum of about 5 kb, and even smaller fragments of 1 to 2 kb are more efficiently amplified. Therefore any variation in the size of the PCR products would have to occur as a result of changes occurring 'between the primers' in a space of a few kb. Regular RFLP probes assay larger portions of the genome, up to 20 kb or more if different restriction enzymes are used. If most RFLP

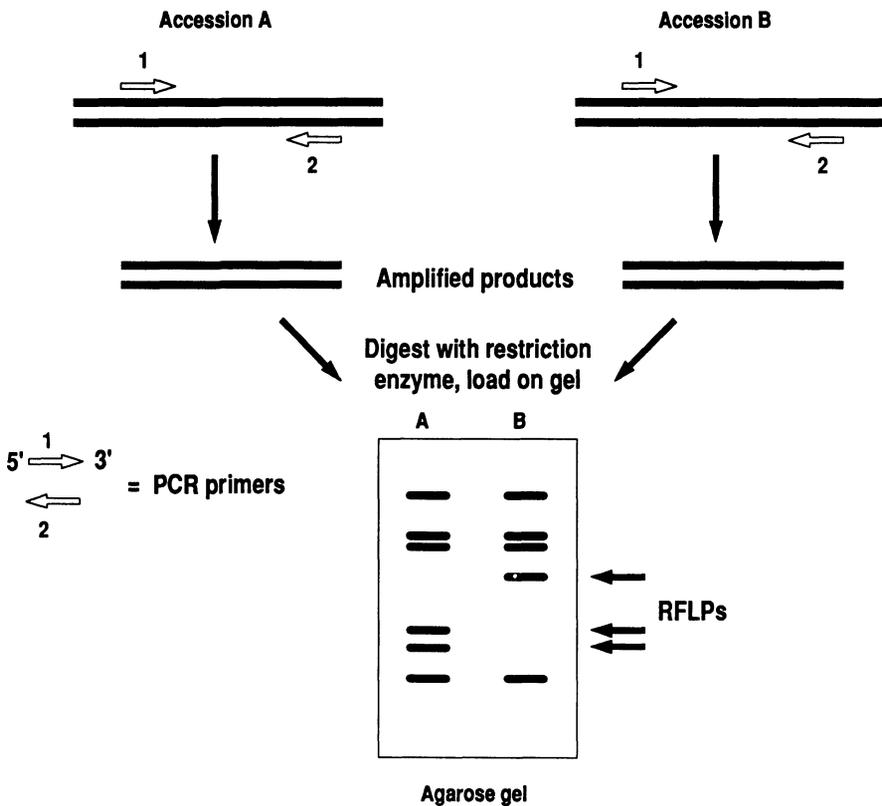


Fig. 4. Detection of genetic differences by restriction enzyme digestion of PCR products of equal length.

variation is caused by genome rearrangements, these are more likely to be found by assaying larger fragments of the genome.

However, even if a length polymorphism is not observed between two PCR products, it is possible to search for single base changes between the primers or small insertion/deletion events which would not be detected by conventional techniques. One way to search for such changes is by restriction enzyme digestion of PCR products. If two parents of a mapping cross produce identically sized products with a given set of PCR primers, the PCR products can be digested with a series of restriction enzymes and the size of the restriction fragments compared by gel electrophoresis (Fig. 4). This is similar to the approach used to gather restriction site information from cloned DNA segments. Since PCR products are rather small, 4-cutter enzymes would be used to increase the likelihood that restriction sites will be present. This approach has been used in studies of fungal population genetics and evolution (Buchko and Klassen 1990; Vilgalys and Gonzalez 1990; Cubeta et al. 1991), but has not been widely employed in plants.

3.1.1. *VNTR loci*

Since PCR with specific primers can only reveal polymorphisms that lie in the amplified area between the primers, one way to increase the utility of PCR based markers is to produce primers that flank genomic regions more likely to show variability than a randomly selected sequence. Such hypervariable regions have been described in a number of organisms and are well characterized in the human genome. These regions consist of tandemly repeated DNA sequences, and the allelic variability is a result of different copy numbers of the tandem repeat being present at different alleles of the same locus. Tandemly repeated DNA sequences have been studied in many organisms. Since the first tandem repeats were isolated as satellite peaks on equilibrium density gradients, tandem repeats are sometimes called 'satellite' DNAs, even when they are not isolated by density gradient centrifugation. When Jeffreys' group characterized the first hypervariable loci, they were found to be tandem repeats of relatively short sequences (15–75 ntp), and they were called 'minisatellites' (Jeffreys et al. 1985a,b). It has also been shown that tandem repeats of even shorter sequences such as $(GT)_n$ or $(CAC)_n$ are common in eucaryotic genomes (Hamada et al. 1984; Tautz and Renz 1984). These loci are termed 'microsatellites' (Litt and Luty 1989), but there is no clear distinction between the two types of loci. A general term for this type of locus is variable number of tandem repeat (VNTR) locus (Nakamura et al. 1987). PCR primers flanking a VNTR locus would be expected to produce maximum length variability and to be the most useful as genetic markers (Fig. 5).

3.1.2. *Minisatellites*

When minisatellite loci were characterized in animal genomes, there was great interest in using these for studies of genetic variability in plants. The quickest way to gather some information was to use the repeat unit of a minisatellite

3.1.3. *Microsatellites*

Microsatellite loci are also found in plants. Tropical tree genomes and maize contain (GT)_n and (AG)_n microsatellites in 10⁴ to 10⁵ copies per genome (Condit and Hubbell 1991). A GGC microsatellite is widely distributed in the rice genome, and several alleles which differ in copy number of the basic repeat are present for one locus which has been characterized (Zhao and Kochert 1992). (GT)_n microsatellites are also found in rice, and when oligonucleotides of (GT)_n are used as hybridization probes on Southern blots of rice genomic DNA, a large number of restriction fragments are detected, and the resultant patterns are useful as DNA fingerprints (X. Zhao and G. Kochert, unpub.). The degree of polymorphism of microsatellite loci is not known in plants, but they are more polymorphic than unique sequence probes in humans (Weber and May 1989; Weber 1990). Thus it seems possible that an entire genetic map could be constructed in any higher plant using one or two classes of microsatellite locus.

It is also possible to use (GT)_n as a single primer in PCR experiments. In this case the region between two adjacent GT microsatellites will produce an amplification product if the microsatellites are in opposite orientation and within a few kb of one another. This protocol also produces patterns useful for rapid screening of genetic variability or DNA fingerprinting (X. Zhao and G. Kochert, unpub.).

3.1.4. *Mapping with VNTR loci*

To enable minisatellite or microsatellite loci to be useful as mapping markers, it is necessary to find efficient ways to locate such loci, to develop methods so that each locus can be studied individually, and to develop a method of rapidly and inexpensively distinguishing between the alleles of a single locus. It should be possible to isolate clones containing VNTR loci from genomic libraries by probing with core sequences from the repeat units of minisatellites. For example, the sequence GGGCAGAXG seems to be found in many of the Jeffreys probes and GGG-GTGGGG is found in many others from animals (Nakamura et al. 1987). Some basic research needs to be done with plants to determine consensus sequences for the plant counterparts. Microsatellite loci can be easily selected from genomic libraries by probing with oligonucleotides such as (CA)_n (Cornall et al. 1991). However, it would be desirable to select microsatellites with a large number of repeats of the basic units because these are reported to be more polymorphic (Weber 1990).

The preferred method for utilizing VNTR markers in mapping studies would be to produce PCR primers complementary to single copy DNA flanking the repeated element. The PCR products could then be easily compared for length variation by electrophoresis. This approach has been demonstrated for both human minisatellites (Jeffreys et al. 1988) and microsatellites (Litt and Luty 1989; Weber and May 1989; Cornall et al. 1991). However, many minisatellite loci are too large to be satisfactorily amplified by PCR, and amounts of PCR product sufficient to detect on stained agarose gels are not formed (Jeffreys et al. 1988). Thus Southern blots and radioactively labelled probes need to be used

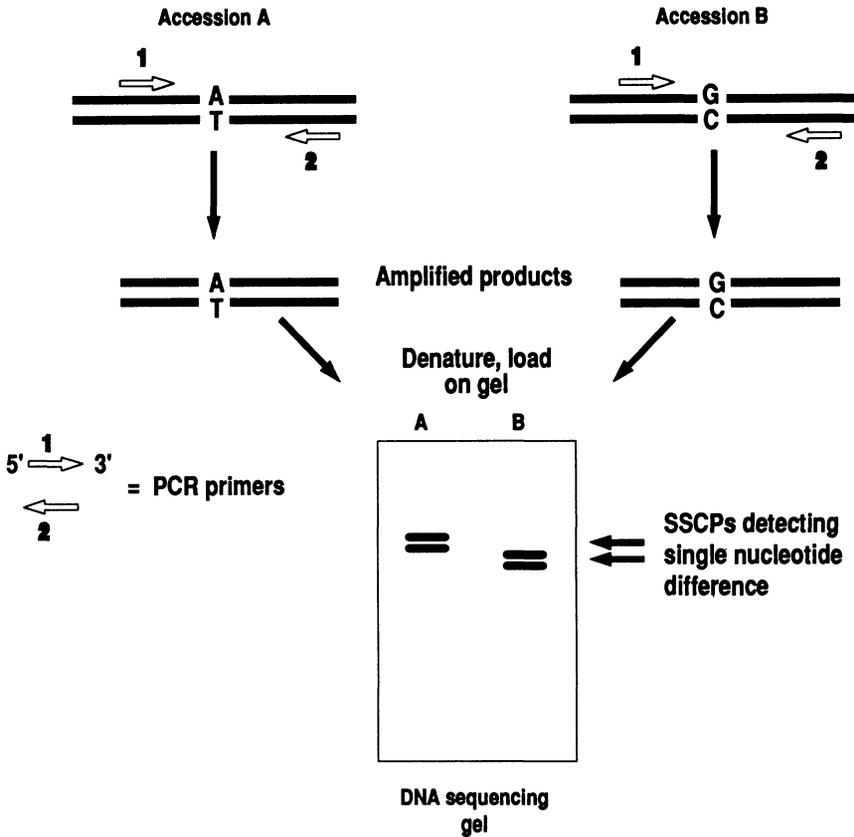


Fig. 6. Detection of single nucleotide differences in PCR products by single strand conformation polymorphisms (SSCPs).

to detect the product. Microsatellite loci are much shorter, and can usually be faithfully amplified by PCR.

To obtain DNA sequence information for areas flanking a VNTR locus, several approaches are possible. Libraries can be constructed of small genomic clones in plasmids. These can be screened for clones containing minisatellite or microsatellite loci, and the sequencing can be performed by extension from 'universal' primers complementary to vector sequences (Cornall et al. 1991). This will work for those clones where the microsatellite portion of the clone is within about 50–200 bases from the end of the cloned restriction fragment, and has enough single copy sequence between the VNTR locus and the vector to enable the design of unique primers. An alternative strategy has been proposed for microsatellite loci (Yuille et al. 1991; Browne and Litt 1992). In this approach the microsatellite sequence itself is used as the basis of a sequencing primer to obtain flanking sequence. One cannot simply use the microsatellite sequence as a sequencing primer, because it could anneal and prime for DNA polymerase in a number of different registers, and the resultant sequence pattern

would be unintelligible. Out-of-register priming can be controlled, however, by constructing a set of primers which contain the microsatellite sequence at the 5' end and each of the other possible non microsatellite motif nucleotides at the 3' end. One of the primers of such a set might be 5'-GTGTGTGTGTGTGTGTGT-AC-3', for example. Such a primer would be forced into a unique register at the boundary of those GT/CA microsatellites where the flanking single copy sequence begins with TG, and not to be able to prime for DNA synthesis if annealed out of register. A set of such primers has been reported to provide flanking sequence for 11 of 12 (CA)_n microsatellites tested (Browne and Litt 1992).

To detect differences in microsatellite alleles it is necessary to detect length differences in PCR products of as little as two nucleotides. Currently, the most sensitive technique is to radioactively label the PCR products and to separate them on a DNA sequencing gel. While this is accurate, it is tedious, time-consuming, and requires the use of isotopes. Simpler separation methods such as direct visualization after staining with ethidium bromide in high-percentage agarose or acrylamide gels resolves the majority (81%) of microsatellite alleles detected in mouse (Cornall et al. 1991), and provides a much quicker and less expensive way of screening for polymorphisms.

3.2. *Single strand conformation polymorphisms*

Another way to detect differences as small as a single nucleotide change in PCR products is to use single strand conformation polymorphisms (SSCPs). This technique relies on the secondary structure being different for single strands derived from PCR products that differ by one or more internal nucleotides (Fig. 6). To detect such differences the PCR products are denatured and electrophoretically separated in neutral acrylamide gels (Orita et al. 1989a,b). Conditions which affect the rate of migration of single-stranded DNA during electrophoresis, such as temperature, buffer concentration, percent acrylamide, and glycerol concentration can be manipulated to achieve separation of most single base changes (Spinardi et al. 1991). Gels in which gradients of denaturants are present (Myers et al. 1985a) can also be used to detect single base changes in PCR products with or without the addition of a 'GC clamp' (Myers et al. 1985b; Riedel et al. 1990), but the gel systems are more complex. SSCP methods have proven useful for detecting single base changes causing mutations in human disease genes (Dean et al. 1990), but they have not been adapted for general use in mapping studies – isotopic labelling and autoradiography are required to detect the products. However, SSCP is a general way of detecting polymorphisms, and their sensitivity is exceeded only by DNA sequencing.

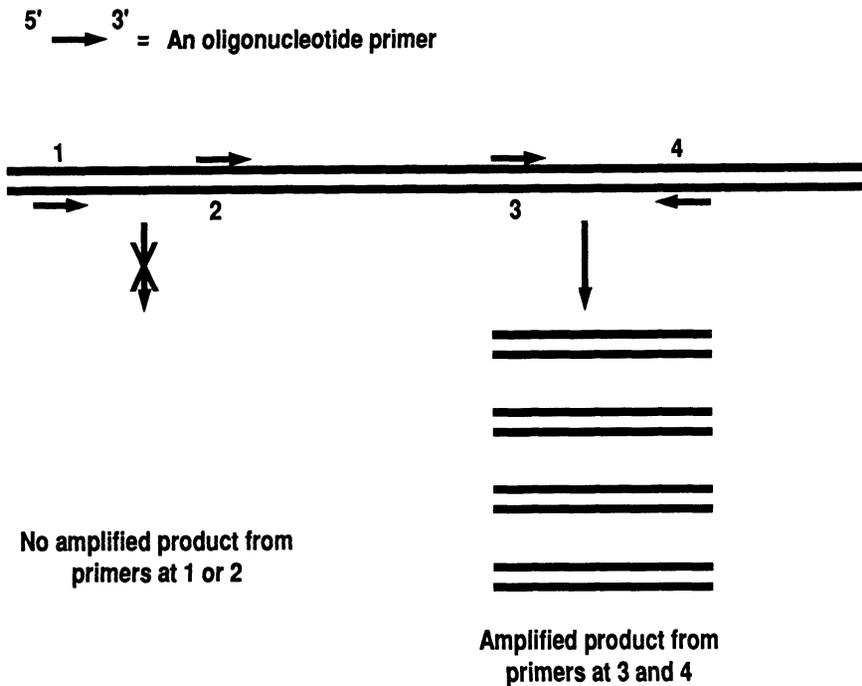


Fig. 7. Production of a PCR product using a single primer. Product will be produced when the sequences in the DNA homologous to the primer are within about 5 kb of one another and have their 3' ends directed towards one another.

3.3. Single primer PCR methods

Conventional PCR is limited in usefulness because of the time and expense required to obtain the DNA sequence information required for the design of primers that will uniquely amplify the desired sequence. To produce the required specificity, primers are usually 20–30 nucleotides in length (a 20 nucleotide primer would be expected to occur at random only once in 4^{20} (10^{12}) nucleotides, and two separate primers are required flanking the sequence to be amplified. If only one 20-nucleotide primer is used, no exponentially amplified product would be expected, because it would be exceedingly unlikely that the same sequence would occur in reverse orientation within a few kb. However, if shorter primers are used it becomes more likely that a single primer will be able to produce an amplified product (Fig. 7). In a large variety of plants and animals it has been shown that single primers 8 to 10 nucleotides in length will produce from one to a few amplification products (Williams et al. 1990; Welsh and McClelland 1990; Caetano-Anolles et al. 1991). Such primers can be synthesized in an oligonucleotide synthesizing facility based on sequences chosen at random; no sequence information is required from the plant to be studied. Choice of which single primers (or RAPD primers) to use is done operationally. Primers are tested in the plant to be studied and those which give reasonably simple and reproducible patterns are selected.

RAPD markers have been enthusiastically received because they provide a quick and simple method to gather information on genetic variability in a wide range of organisms. In addition the analyses can be performed with only a small amount of DNA. RAPD markers are being widely used in surveys of genetic variability (Arnold et al. 1991; Welsh et al. 1991; Welsh and McClelland 1991), and the information gained can be used for taxonomic purposes (Halward et al. 1992) or for the detection of genome-specific markers (Quiros et al. 1991; Klein-Lankhorst et al. 1991). Their potential as genetic markers for map construction is also being exploited (Carlson et al. 1991; Echt et al. 1992), and maps consisting largely or exclusively of RAPD markers have been constructed in *Arabidopsis* (Reiter et al. 1992), pine (R. Sederoff, pers. comm.) and *Helianthus*. (L. Rieseberg, pers. comm.). However, when used as genetic markers for map construction, RAPD markers have some limitations which offset some of their undoubted advantages. First, most RAPD polymorphisms are inherited as dominant-recessive characters. The most common RAPD polymorphism is the presence of a band in one parent of a mapping cross and the absence of that band in the other parent (Williams et al. 1990; Reiter et al. 1992). Since RAPD primers are relatively short, a mismatch of even a single nucleotide can often prevent the primer from annealing, and this is thought to be the most common mechanism for RAPD polymorphisms. The same pattern could, of course, be generated by deletion of one of the primer sites or an insertion between the primer sites that would move them too far apart to be able to produce an amplification product. In any case, segregating progeny can usually only be scored for the presence or absence of a given RAPD marker, and the heterozygote cannot be distinguished from the homozygote because the PCR reaction usually cannot be quantitated. Thus in an F_2 population, only two classes of progeny can be distinguished: those which do not exhibit the band in question (homozygous for the absence of the band), and those which have the band (homozygotes for the presence of the band plus heterozygotes). In an F_2 these two classes would be expected to segregate in a 3 present:1 not present ratio. This causes a loss of information relative to markers which show codominance, such as conventional RFLP markers. Backcross, recombinant inbred, and doubled haploid populations do not suffer this loss of mapping information (they are equally informative with codominant and dominant-recessive markers), because the complete information available from a backcross can be obtained from scoring the presence or absence of a polymorphic marker (Reiter et al. 1992).

The presence of an elaborated gametophyte stage in gymnosperms has also been exploited for genetic map construction using RAPD markers (R. Sederoff, pers. comm.). Gametophyte tissue is, of course, haploid, and each female gametophyte produced on an F_1 hybrid plant is derived from a single meiotic product. DNA samples derived from a set of gametophytes from a single F_1 plant thus constitute a mapping population which is equivalent to a backcross population without the necessity of performing the backcrosses and without the complication of recurrent parent DNA being present. Presence or absence of

RAPD bands can be scored in these gametophyte-derived DNAs, and genetic maps can be rapidly constructed. This system could be extended to angiosperms in those cases where microspores can be cultivated to give haploid callus.

Various other sequences can be used as single primers for PCR, such as t-RNA consensus sequences (Welsh and McClelland 1991). Microsatellite sequences, such as (GT)_n can also be used as single primers for DNA fingerprinting, but the patterns produced are too complex to be useful in the construction of genetic maps (X. Zhao and G. Kochert, unpub.).

3.4. *Locating markers in defined chromosomal segments*

In the application of molecular maps to plant breeding applications it is often desirable to 'tag' an agronomically valuable gene with closely linked molecular markers. One way to accomplish this is to map the trait relative to a set of molecular markers in a segregating population. Since the trait of interest is not usually found in the mapping population originally used to construct the molecular map, it is necessary to rescore the segregation of some of the markers in a different population, one that is segregating for the trait. If there is no prior knowledge from conventional mapping about the location of the gene, molecular markers have to be rescored systematically in the new population until one is found which is linked.

More efficient ways of obtaining linked markers utilize near-isogenic lines (NILs). NILs have been produced for a variety of traits in several crop plants by repeated cycles of backcrossing and selection. Ideally, they consist of a single chromosome segment containing the trait of interest. If there is sufficient polymorphism for molecular markers between the parent lines used to make the NIL, molecular markers can be rapidly screened to locate those which are in the introgressed segment and thus polymorphic between the recurrent parent and the NIL. RFLP markers (Young et al. 1988; Muehlbauer et al. 1991; Yu et al. 1991) or RAPD markers can be used (Martin et al. 1991; Paran et al. 1991). Many RAPDs can be screened in a short time, and the increased cost may be justifiable in some cases (Martin et al. 1991). No mapping or use of segregating populations is theoretically necessary to locate linked markers. However, NILs so far investigated have been found to contain segments of introgressed DNA other than the one selected in the breeding program, and it is necessary to check the putative positive clones by segregation analysis (Young et al. 1988).

Pooling of DNA samples can also be used to locate molecular markers in defined chromosomal segments or those which are closely linked to genes of interest. This is, in effect, a way of creating isogenic lines *in vitro*. This approach was used by Arnheim et al. (1985) to locate RFLP markers closely linked to human disease genes. Micheltore et al. (1991) called the technique 'bulked segregant analysis'. F₂ plants from a lettuce population segregating for resistance to downy mildew were separated into two groups: those which were homozygous resistant and those which were homozygous for sensitivity (heterozygotes previously identified by progeny tests were excluded). DNA was

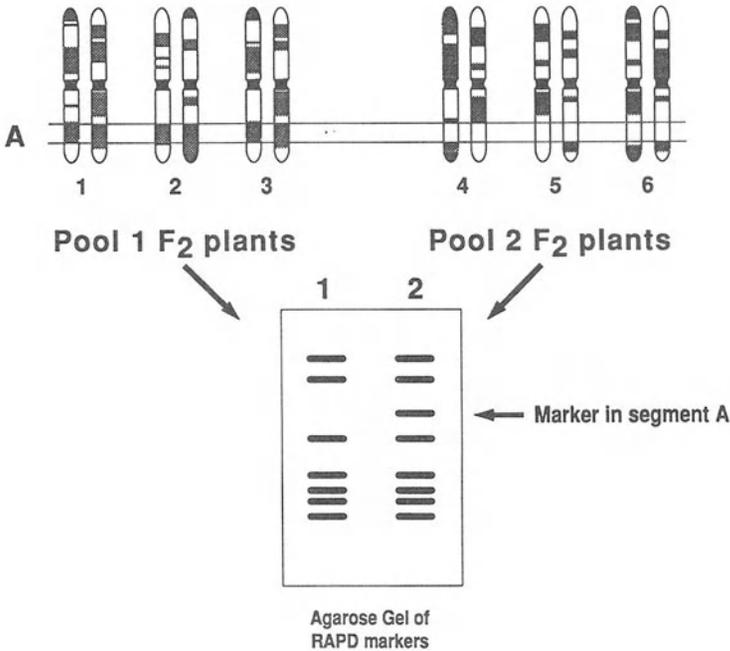


Fig. 8. Detection of RAPD polymorphisms in a defined chromosome segment by pooling DNA from a segregating population of F₂ plants.

then combined from several plants from each group to form two DNA pools. RAPD primers were then used to amplify DNA from each of the two pools. The pooled DNA samples should be homozygous and different from each other in the chromosomal area around the downy mildew resistance locus, but should contain DNA from both parents for all other chromosome segments. Thus any polymorphisms seen in the RAPD analysis should represent markers in the area near the resistance gene. Three RAPD polymorphisms were identified and all were linked to the resistance gene. Giovannoni et al. (1991) described a related method in which the DNA pools are constructed on the basis of mapped molecular markers (Fig. 8). A target segment of the genome is selected which has a gene of interest or represents a segment where more markers are desired. Two DNA pools are constructed: one pool contains bulked DNA from F₂ plants homozygous for parent 1 markers flanking the segment of interest; the other pool is from F₂ plants homozygous for markers from the other parent. Again, since no selection has been carried out for other segments of the genome, each pool should contain genome segments from both parents for segments other than the target segment. Homozygous pools can also be compared to heterozygous pools. Polymorphisms identified by RAPD analysis should then represent markers located in the target segment. Giovannoni et al. (1991) described the utility of this method in tomato, and compared their method to bulked segregant analysis. These methods have great potential for the rapid

tagging of agronomically important genes, and they will relieve the need to construct NILs in many cases. Individual recombinant inbred lines or doubled haploid lines can be pooled in a similar way (Reiter et al. 1992).

4. Future directions

It is evident that the development of DNA markers has revolutionized the construction of genetic maps in plants and the utilization of genetic maps in studies of plant evolution, systematics, and practical applications such as plant breeding. DNA markers allow direct access to any part of a plant genome, and they liberate researchers from having to deal with plant genes through the fog of phenotype, many steps away from the gene itself. Technology for the utilization of DNA markers is evolving rapidly at the present time, and further advances are sure to occur soon. Some of these will involve making the process of developing and utilizing DNA markers technically simpler, less expensive, and more capable of automation. To be practical on a large scale for plant breeding applications, particularly in developing countries, detection procedures for DNA markers need to be developed which do not require the use of radioactive isotopes, Southern blots, DNA sequencing gels and the like. PCR based methods such as RAPD analysis seem to provide part of the answer, but these procedures are still very expensive because of high reagent costs. Simplified DNA analyses seem to be possible with PCR, and even tissue squashes may suffice for DNA isolation (Langridge et al. 1991).

The utility of maps of molecular markers will continue to increase. The ability to rapidly construct genetic maps has made possible applications that were unthinkable using conventional mapping techniques. Comparative mapping of different crop plants (Bonierbale et al. 1988; Tanksley et al. 1988a) will provide useful information about the location of important genes, because it is likely that there will be enough conservation of syntenic blocks so that genes located in one plant will have the same flanking markers in another plant. Comparative mapping of crop plants and their wild relatives will be a valuable tool for phylogenetic analysis, as well as being useful in introgression studies.

One of the most important uses of molecular maps will be for gene cloning. Transformation systems for many crop plants are being developed, but one of the main limitations of this technique is a lack of useful genes for transformation. Most genes for valuable agronomic traits such as disease and insect resistance have not been cloned, and thus are not available for transformation. In addition, there is so little known about the mode of action of genes of agronomic importance, it seems unlikely that many of them could be cloned by conventional cloning protocols that depend on having the product of the gene to use as a probe to screen genomic or cDNA libraries. In most cases nothing is known about the product of these genes, there is no generally useful and efficient way of determining the product, and no homologous cloned genes are available from other organisms to serve as heterologous probes.

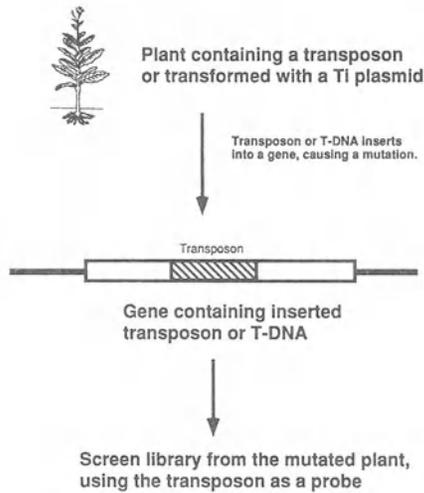


Fig. 9. Insertion of a transposon or T-DNA to provide a probe for gene cloning.

There are two general ways to clone a gene without any knowledge of its product: insertion mutagenesis and map-based cloning. Insertion mutagenesis tags a gene by inserting a known sequence, which can be used as a probe to screen a library, into the gene. This known sequence can be a transposon such as *Ac*, which is naturally present (Fedoroff 1991), or can be transformed into the plant of interest (Taylor et al. 1989; Izawa et al. 1991; Belzile and Yoder 1992) or a sequence which is introduced into the plant by transformation, such as T-DNA from *Agrobacterium tumefaciens* (Feldmann et al. 1989; Koncz et al. 1990; Errampalli et al. 1991; Feldmann 1991; Walden et al. 1991). Plants are then screened for a mutation which is indicative of the insertion having gone into the gene of interest. The transposon or T-DNA sequence is then used as a probe to screen a genomic library made from the mutated plant (Fig. 9). This tagging approach has worked very well in some cases, but it has obvious limitations for plants which cannot be transformed with high efficiency, do not have natural transposon systems, and for genes that do not exhibit an easily screened phenotype when mutated.

The most general way of cloning genes is map-based cloning, because any gene which can be genetically mapped can theoretically be cloned by this method. The basic idea is to locate molecular markers which closely flank the gene of interest, then to use these markers as probes to initiate the isolation of overlapping adjacent clones from a genomic library. Eventually, by such chromosome 'walking' a DNA fragment containing the desired gene will be obtained (Fig. 10). The ability to carry out map-based cloning is, of course, dependent on having inexpensive and efficient methods for the construction of molecular maps, but molecular maps can now be constructed rapidly for nearly all plants, and the limiting time factor is often the production of appropriate mapping populations rather than the production of the map itself. Methods also

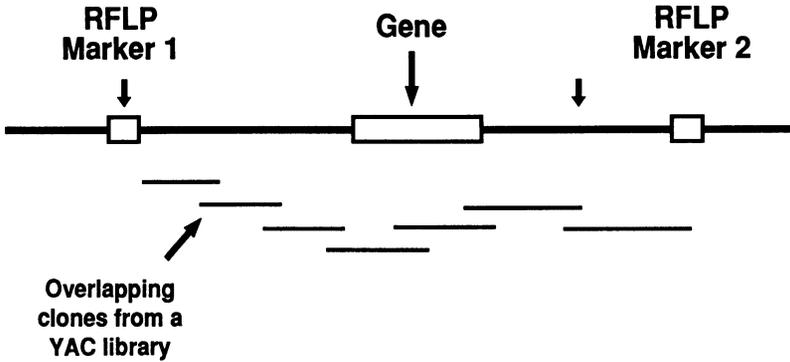


Fig. 10. Selection of overlapping clones to clone a gene by chromosome walking.

exist in plants for finding markers which are closely linked to genes of interest (see above), but the current limitation is in the transition from genetic mapping to physical mapping of the appropriate chromosome segment, which is required to identify and clone the desired gene. Since there is no direct correspondence between genetic distances as determined by genetic mapping and physical distances in nucleotide pairs (Meagher et al. 1988; Young 1990), it is difficult to predict for any one case just how much DNA will be present between flanking markers and the desired gene. Clearly, however, it will be necessary to work with large DNA segments to efficiently carry out map-based cloning. Electrophoretic methods for the analysis of large DNA fragments have been developed, and are collectively called pulsed-field gel methods (Cantor et al. 1988). Protocols have been worked out for the isolation and analysis of large DNA fragments from rice (Sobral et al. 1992), wheat (Cheung and Gale 1990), tomato (Ganal et al. 1989; Ganal and Tanksley 1989), and soybean (Honeycutt et al. 1992) by pulsed-field electrophoresis, and large DNA fragments can be cloned into vectors such as yeast artificial chromosomes (YACs) (Schlessinger 1990). YAC libraries are being developed for several plants (Ward and Jen 1990; Grill and Somerville 1991; Hwang et al. 1991; Dunford and Rogner 1991), and this technology will benefit from advances in the various plant and animal genome projects currently underway. Automated, inexpensive, DNA sequencing protocols would particularly benefit map-based cloning efforts.

Cloning and transformation techniques are now feasible only for traits controlled by single genes, but most agronomically important traits are polygenic, and several genes or quantitative trait loci (QTLs) are required for full expression. Molecular maps of DNA markers have made it possible to begin analyzing QTLs to determine how many might be involved in each trait and the location of the chromosomal segment containing them (Paterson et al. 1988, 1990, 1991; Tanksley and Hewitt 1988; Lander and Botstein 1989; Knapp et al. 1990). However, locating, characterizing and eventually cloning QTLs adds another layer of complexity to the study of single gene traits. Nevertheless, DNA markers are powerful tools for a new assault on basic problems which

have always been faced by those who study plant development and evolution, and those who strive to produce improved plants for human use

5. References

- Allard, R W (1956) Formulas and tables to facilitate the calculation of recombination values in heredity *Hilgardia* 24 235–278
- Apuya, N R , Frazier, B L , Keim, P , Roth, E J and Lark, K G (1988) Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* (L) Merrill *Theor Appl Genet* 75 889–991
- Arnheim, N , Strange, C and Erlich, H (1985) Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease studies of the *HLAII* loci *Proc Natl Acad Sci U S A* 82 6970–6974
- Arnold, M L , Buckner, C M and Robinson, J J (1991) Pollen-mediated introgression and hybrid speciation in Louisiana irises *Proc Natl Acad Sci U S A* 88 1398–1402
- Arumuganathan, K and Earle, E D (1991) Nuclear DNA content of some important plant species *Plant Mol Biol Rep* 9 208–218
- Belzile, F and Yoder, J I (1992) Pattern of somatic transposition in a high copy Ac tomato line *Plant J* 2 173–179
- Bennett, M D and Smith, J B (1976) Nuclear DNA amounts in angiosperms *Phil Trans R Soc London Ser B* 274 227–274
- Bennett, M D and Smith, J B (1991) Nuclear DNA amounts in angiosperms *Phil Trans R Soc London Ser B* 334 309–345
- Bennett, M D , Smith, J B and Heslop-Harrison, J S (1982) Nuclear DNA amounts in angiosperms *Phil Trans R Soc London Ser B* 216 179–199
- Bonierbale, M W , Plaisted, R L and Tanksley, S D (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato *Genetics* 120 1095–1103
- Botstein, D , White, R L , Skolnick, M and Davis, R W (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms *Am J Hum Genet* 32 314–331
- Brown, P T H , Gobel, E and Lorz, H (1991) RFLP analysis of *Zea mays* callus cultures and their regenerated plants *Theor Appl Genet* 81 227–232
- Browne, D L and Litt, M (1992) Characterization of (CA)_n microsatellites with degenerate sequencing primers *Nucl Acids Res* 20 141
- Brummer, E C , Kochert, G and Bouton, J H (1991) RFLP variation in diploid and tetraploid alfalfa *Theor Appl Genet* 83 89–96
- Buchko, J and Klassen, G R (1990) Detection of length heterogeneity in the ribosomal DNA of *Pythium ultimum* by PCR amplification of the intergenic region *Curr Genet* 18 203–205
- Burr, B and Burr, F A (1991) Recombinant inbreds for molecular mapping in maize theoretical and practical considerations *Trends Genet* 7 55–60
- Burr, B , Burr, F A , Thompson, K H , Albertson, M C and Stuber, C W (1988) Gene mapping with recombinant inbreds in maize *Genetics* 118 519–526
- Caetano-Anolles, G , Bassam, B J and Gresshoff, P M (1991) DNA amplification fingerprinting using very short arbitrary oligonucleotide primers *Bio/Technology* 9 553–557
- Cantor, C R , Smith, C L and Mathew, M K (1988) Pulsed-field gel electrophoresis of very large DNA molecules *Ann Rev Biophys Biophys Chem* 17 287–304
- Carlson, J E , Tulsieram, L K , Glaubitz, J C , Luk, V W K , Kauffeldt, C and Rutledge, R (1991) Segregation of random amplified DNA markers in F₁ progeny of conifers *Theor Appl Genet* 83 194–200
- Cheung, W Y and Gale, M D (1990) The isolation of high molecular weight DNA from wheat, barley and rye for analysis by pulse-field gel electrophoresis *Plant Mol Biol* 14 881–888

- Condit, R and Hubbell, S P (1991) Abundance and DNA sequence of two-base repeat regions in tropical tree genomes *Genome* 34 66–71
- Cornall, R J , Aitman, T J , Hearne, C M and Todd, J A (1991) The generation of a library of PCR-analyzed microsatellite variants for genetic mapping of the mouse genome *Genomics* 10 874–881
- Crow, J F and Dove, W F (1988) A diamond anniversary the first chromosome map *Genetics* 118 389–391
- Cubeta, M A , Echandi, E , Abernethy, T and Vilgalys, R (1991) Characterization of anastomosis groups of binucleate *Rhizoctonia* species using restriction analysis of an amplified ribosomal RNA gene *Phytopathology* 81 1395–1400
- De Vicente, M C and Tanksley, S D (1991) Genome-wide reduction in recombination of backcross progeny derived from male versus female gametes in an interspecific cross of tomato *Theor Appl Genet* 83 173–178
- D'Ovidio, R , Tanzarella, O A and Porceddu, E (1990) Rapid and efficient detection of genetic polymorphism in wheat through amplification by polymerase chain reaction *Plant Mol Biol* 15 169–171
- Dallas, J F (1988) Detection of DNA 'fingerprints' of cultivated rice by hybridization with a human minisatellite DNA probe *Proc Natl Acad Sci U S A* 85 6831–6835
- Dean, M , White, M B , Amos, J , Gerrard, B , Stewart, C , Khaw, K -T and Leppert, M (1990) Multiple mutations in highly conserved residues are found in mildly affected cystic fibrosis patients *Cell* 61 863–870
- Donis-Keller, H , Green, P , Helms, C , Cartinhour, S , Weiffenbach, B , Stephens, K , Keith, T P , Bowden, D W , Smith, D R , Lander, E S , Botstein, D , Akots, G , Rediker, K S , Gravius, T , Brown, V A , Rising, M B , Parker, C , Powers, J A , Watt, D E , Kauffman, E R , Bricker, A , Phipps, P , Muller-Kahle, H , Fulton, T R , Ng, S , Schumm, J W , Braman, J C , Knowlton, R G , Barker, D F , Crooks, S M , Lincoln, S E , Daly, M J and Abrahamson, J (1987) A genetic linkage map of the human genome *Cell* 51 319–337
- Dunford, R and Rogner, U C (1991) Construction of yeast artificial chromosomes containing barley DNA and the identification of clones carrying copies of the repeated element BIS-1 *Hereditas* 115 133–138
- Echt, C S , Erdahl, L A and McCoy, T J (1992) Genetic segregation of random amplified polymorphic DNA in diploid cultivated alfalfa *Genome* 35 84–87
- Errampalli, D , Patton, D , Castle, L , Mickelson, L , Hansen, K , Schnall, J , Feldmann, K and Meinke, D (1991) Embryonic lethals and T-DNA insertional mutagenesis in *Arabidopsis* *Plant Cell* 3 149–157
- Fedoroff, N V (1991) Maize transposable elements *Perspect Biol Med* 35 2–19
- Feinberg, A P and Vogelstein, B (1984) A technique for radiolabelling DNA restriction fragments to a high specific activity *Anal Biochem* 132 6–13
- Feldmann, K A (1991) T-DNA insertion mutagenesis in *Arabidopsis* mutational spectrum *Plant J* 1 71–82
- Feldmann, K A , Marks, M D , Christianson, M L and Quatrano, R S (1989) A dwarf mutant of *Arabidopsis* generated by T-DNA insertion mutagenesis *Science* 243 1351–1354
- Figdore, S S , Kennard, W C , Song, K M , Slocum, M K and Osborn, T C (1988) Assessment of the degree of restriction fragment length polymorphism in *Brassica* *Theor Appl Genet* 75 833–840
- Friar, E and Kochert, G (1991) Bamboo germplasm screening with nuclear restriction fragment length polymorphisms *Theor Appl Genet* 82 697–703
- Ganal, M W and Tanksley, S D (1989) Analysis of tomato DNA by pulsed field gel electrophoresis *Plant Mol Biol Rep* 7 17–28
- Ganal, M W , Young, N D and Tanksley, S D (1989) Pulsed field gel electrophoresis and physical mapping of large DNA fragments in the *Tm-2a* region of chromosome 9 in tomato *Mol Gen Genet* 215 395–400
- Giovannoni, J J , Wing, R A , Ganal, M W and Tanksley, S D (1991) Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations *Nucl Acids Res* 19 6553–6558

- Grill, E and Somerville, C (1991) Construction and characterization of a yeast artificial chromosome library of *Arabidopsis* which is suitable for chromosome walking Mol Gen Genet 226 484-490
- Halward, T, Stalker, T, Larue, E and Kochert, G (1992) Use of single-primer DNA amplifications in genetic studies of peanut (*Arachis hypogaea* L) Plant Mol Biol 18 315-325
- Hamada, H, Petrino, M G, Kakunaga, T, Seidman, M and Stollar, B D (1984) Characterization of genomic poly(dT-dG)-poly(dC-dA) sequences structure, organization and conformation Mol Cell Biol 4 2610-2621
- Helentjaris, T (1987) A genetic linkage map for maize based on RFLPs Trends Genet 3 217-221
- Helentjaris, T, King, G, Slocum, M, Sidenstrang, C and Wegmen, S (1985) Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding Plant Mol Biol 5 109-118
- Heun, M, Kennedy, A E, Anderson, J A, Lapitan, N L V, Sorrells M E and Tanksley, S D (1991) Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*) Genome 34 437-447
- Honeycutt, R J, Sobral, B W S, McClelland, M and Atherly, A G (1992) Analysis of large DNA from soybean (*Glycine max* L Merr) by pulsed-field gel electrophoresis Plant J 2 133-135
- Hwang, I, Kohchi, T, Hauge, B M, Goodman, H M, Schmidt, R, Cnops, G, Dean, C, Gibson, S, Iba, K, Lemieux, B, Arondel, V, Danhoff, L and Somerville, C (1991) Identification and map position of YAC clones comprising one-third of the *Arabidopsis* genome Plant J 1 367-374
- Ishii, T, Panaud, O, Brar, D S and Khush, G S (1990) Use of non-radioactive digoxigenin-labelled DNA probes for RFLP analysis in rice Plant Mol Biol Rep 8 167-171
- Izawa, T, Miyazaki, C, Yamamoto, M, Terada, R, Iida, S and Shimamoto, K (1991) Introduction and transposition of the maize transposable element *Ac* in rice (*Oryza sativa* L) Mol Gen Genet 227 391-396
- Jeffreys, A J, Wilson, V and Thein, S L (1985a) Hypervariable 'minisatellite' regions in human DNA Nature 314 67-73
- Jeffreys, A J, Wilson, V and Thein, S L (1985b) Individual-specific 'fingerprints' of human DNA Nature 314 76-79
- Jeffreys, A J, Wilson, V, Neumann, R and Keyte, J (1988) Amplification of human minisatellites by the polymerase chain reaction towards DNA fingerprinting of single cells Nucl Acids Res 16 10953-10972
- Johnson, B S and Th'ng, P H (1991) DNA fingerprinting with ³⁵S nucleotides Nucl Acids Res 18 7459-7460
- Kasha, K J and Sadasivaiah, R S (1971) Genome relationships between *Hordeum vulgare* L and *H bulbosum* L Chromosoma 35 264-287
- Klein-Lankhorst, R M, Vermunt, A, Weide, R, Liharska, T and Zabel, P (1991) Isolation of molecular markers for tomato (*L. esculentum*) using random amplified polymorphic DNA (RAPD) Theor Appl Genet 83 108-114
- Knapp, S J, Bridges, W C and Birkes, D (1990) Mapping quantitative trait loci using molecular marker linkage maps Theor Appl Genet 79 583-592
- Kochert, G, Halward, T, Branch, W D and Simpson, C E (1991) RFLP variability in peanut (*Arachis hypogaea*) cultivars and wild species Theor Appl Genet 81 565-570
- Koncz, C, Mayerhofer, R, Konczkalman, Z, Nawrath, C, Reiss, B, Redei, G P and Schell, J (1990) Isolation of a gene encoding a novel chloroplast protein by T-DNA tagging in *Arabidopsis thaliana* EMBO J 9 1337-1346
- Lander, E S and Botstein, D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps Genetics 121 185-199
- Landry, B S and Michelmore, R W (1987) Methods and applications of restriction fragment length polymorphism analysis to plants In T Kosuge and A Hollaender (eds), Tailoring Genes for Crop Improvement, pp 25-44 Plenum Publishing Corp, New York
- Landry, B S, Kesseli, R, Leung, H and Michelmore, R W (1987) Comparison of restriction endonucleases and sources of probes for their efficiency in detecting restriction fragment length polymorphisms in lettuce (*Lactuca sativa* L) Theor Appl Genet 74 646-653

- Langridge, U., Schwall, M. and Langridge, P. (1991) Squashes of plant tissue as substrate for PCR. *Nucl. Acids Res.* 19: 6954.
- Litt, M. and Luty, J.A. (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 44: 397-401.
- Liu, Y.G., Mori, N. and Tsunewaki, K. (1990) Restriction fragment length polymorphism (RFLP) analysis in wheat. I. Genomic DNA library construction and RFLP analysis in common wheat. *Jpn. J. Genet.* 65: 367-380.
- Martin, G.B., Williams, J.G.K. and Tanksley, S.D. (1991) Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. *Proc. Natl. Acad. Sci. U.S.A.* 88: 2336-2340.
- McCouch, S.R., Kochert, G., Yu, Z.H., Wang, Z.Y., Khush, G.S., Coffman, W.R. and Tanksley, S.D. (1988) Molecular mapping of rice chromosomes. *Theor. Appl. Genet.* 76: 815-829.
- Meagher, R.B., McLean, M.D. and Arnold, J. (1988) Recombination within a subclass of restriction fragment length polymorphisms may help link classical and molecular genetics. *Genetics* 120: 809-818.
- Michelmore, R.W., Paran, I. and Kesseli, R.V. (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. U.S.A.* 88: 9828-9832.
- Miller, J.C. and Tanksley, S.D. (1990) Effect of different restriction enzymes, probe source, and probe length on detecting restriction fragment length polymorphism in tomato. *Theor. Appl. Genet.* 80: 385-389.
- Muehlbauer, G.J., Staswick, P.E., Specht, J.E., Graef, G.L., Shoemaker, R.C. and Keim, P. (1991) RFLP mapping using near-isogenic lines in the soybean [*Glycine max* (L.) Merr]. *Theor. Appl. Genet.* 81: 189-198.
- Muller, E., Brown, P.T.H., Hartke, S. and Lorz, H. (1990) DNA variation in tissue-culture-derived rice plants. *Theor. Appl. Genet.* 80: 673-679.
- Murray, M. and Thompson, W.F. (1980) Rapid isolation of high molecular weight plant DNA. *Nucl. Acids Res.* 8: 4321-4325.
- Myers, R.M., Lumelsky, N., Lerman, L.S. and Maniatis, T. (1985a) Detection of single base substitutions in total genomic DNA. *Nature* 313: 495-498.
- Myers, R.M., Fischer, S.G., Maniatis, T. and Lerman, L.S. (1985b) Modification of the melting properties of duplex DNA by attachment of a GC-rich DNA sequence as determined by denaturing gradient gel electrophoresis. *Nucl. Acids Res.* 13: 3111-3129.
- Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235: 1616-1622.
- Nybom, H. (1990) Genetic variation in ornamental apple trees and their seedlings (*Malus*, Rosaceae) revealed by DNA fingerprinting with the M13 repeat probe. *Hereditas* 113: 17-28.
- Nybom, H., Rogstad, S.H. and Schaal, B.A. (1990) Genetic variation detected by use of the M13 DNA fingerprint probe in *Malus*, *Prunus*, and *Rubus* (Rosaceae). *Theor. Appl. Genet.* 79: 153-156.
- Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T. (1989a) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. U.S.A.* 86: 2766-2770.
- Orita, M., Suzuki, Y., Sekiya, T. and Hayashi, K. (1989b) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* 5: 874-879.
- Paran, I., Kesseli, R. and Michelmore, R. (1991) Identification of restriction fragment length polymorphism and random amplified polymorphic DNA markers linked to downy mildew resistance genes in lettuce, using near-isogenic lines. *Genome* 34: 1021-1027.
- Paterson, A.H., Lander, E.S., Hewitt, J.D., Peterson, S., Lincoln, S.E. and Tanksley, S.D. (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335: 721-726.
- Paterson, A.H., Deverna, J.W., Lanini, B. and Tanksley, S.D. (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. *Genetics* 124: 735-742.

- Paterson, A H , Damon, S , Hewitt, J D , Zamir, D , Rabinowitch, H D , Lincoln, S E , Lander, E S and Tanksley, S D (1991) Mendelian factors underlying quantitative traits in tomato comparison across species, generations, and environments *Genetics* 127 181-197
- Powell, W , Thomas, W T B , Thompson, D M , Swanston, J S and Waugh, R (1992) Association between rDNA alleles and quantitative traits in doubled haploid populations of barley *Genetics* 130 187-194
- Quiros, C F , Hu, J , This, P , Chevre, A M and Delseny, M (1991) Development and chromosomal localization of genome-specific markers by polymerase chain reaction in *Brassica* *Theor Appl Genet* 82 627-632
- Reiter, R S , Williams, J G K , Feldmann, K A , Rafalski, J A , Tingey, S V and Scolnik, P A (1992) Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs *Proc Natl Acad Sci U S A* 89 1477-1481
- Riedel, G E , Swanberg, S L , Kuranda, K D , Marquette, K , Lapan, P , Bledsoe, P , Kennedy, A and Lin, B Y (1990) Denaturing gradient gel electrophoresis identifies genomic DNA polymorphism with high frequency in maize *Theor Appl Genet* 80 1-10
- Rigby, P , Dieckmann, M , Rhodes, C and Berg, P (1977) Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick-translation with DNA polymerase I *J Mol Biol* 113 237-251
- Ritter, E , Gebhardt, C and Salamini, F (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents *Genetics* 125 645-654
- Rivard, S R , Cappadocia, M , Vincent, G , Brisson, N and Landry, B S (1989) Restriction fragment length polymorphism (RFLP) analyses of plants produced by *in vitro* anther culture of *Solanum chacoense* Bitt *Theor Appl Genet* 78 49-56
- Rogers, S O and Bendich, A J (1988) Extraction of DNA from plant tissues. In S B Gelvin, R A Schilperoort and D P S Verma (eds), *Plant Molecular Biology Manual*, pp 1-11 Kluwer Academic Publishers, Boston
- Rogstad, S H , Patton, J C , II and Schaal, B A (1988) M13 repeat probe detects DNA minisatellite-like sequences in gymnosperms and angiosperms *Proc Natl Acad Sci U S A* 85 9176-9178
- Rogstad, S H , Wolff, K and Schaal, B A (1991) Geographical variation in *Asimina triloba* Dunal (Annonaceae) revealed by the M13 DNA fingerprinting probe *Am J Bot* 78 1391-1396
- Roth, E J , Frazier, B L , Apuya, N R and Lark, K G (1989) Genetic variation in an inbred plant variation in tissue cultures of soybean [*Glycine max* (L) Merrill] *Genetics* 121 359-368
- Saghar-Marroof, M A , Soliman, K M , Jorgensen, R A and Allard, R W (1984) Ribosomal DNA spacer-length polymorphisms in barley Mendelian inheritance, chromosomal location, and population dynamics *Proc Natl Acad Sci U S A* 81 8014-8018
- Schlessinger, D (1990) Yeast artificial chromosomes tools for mapping and analysis of complex genomes *Trends Genet* 6 248
- Schlotterer, C and Tautz, D (1992) Slippage synthesis of simple sequence DNA *Nucl Acids Res* 20 211-215
- Sobral, W S , Honeycutt, R J , Atherly, A G and McClelland, M (1992) Analysis of rice (*Oryza sativa* L) genome using pulsed-field gel electrophoresis and rare-cutting restriction endonucleases *Plant Mol Biol Rep* 8 253-275
- Southern, E M (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis *J Mol Biol* 98 503-517
- Spinardi, L , Mazars, R and Theillet, C (1991) Protocols for an improved detection of point mutations by SSCP *Nucl Acids Res* 19 4009
- Tanksley, S D and Hewitt, J (1988) Use of molecular markers in breeding for soluble solids content in tomato a re-examination *Theor Appl Genet* 75 811-823
- Tanksley, S D , Bernatzky, R , Lapitan, N L and Prince, J P (1988a) Conservation of gene repertoire but not gene order in pepper and tomato *Proc Natl Acad Sci U S A* 85 6419-6423
- Tanksley, S D , Miller, J , Paterson, A and Bernatzky, R (1988b) Molecular mapping of plant chromosomes. In J P Gustafson and R Appels (eds), *Chromosome Structure and Function*, pp 157-173 Plenum Publishing Corp , New York

- Tautz, D and Renz, M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes *Nucl Acids Res* 12 4127–4138
- Taylor, B H , Finnegan, E J , Dennis, E S and Peacock, W J (1989) The maize transposable element *Ac* excises in progeny of transformed tobacco *Plant Mol Biol* 13 109–118
- Vahala, T , Eriksson, T and Engstrom, P (1991) Genetic variability in basket willow (*Salix viminalis*) detected by hybridization to a bacteriophage-M13 DNA probe *Hereditas* 115 153–161
- Vassart, G , Georges, M , Monsieur, R , Brocas, H , Lequarre, A S and Christophe, D (1987) A sequence in M13 detects hypervariable minisatellites in human and animal DNA *Science* 235 683–684
- Vilgalys, R and Gonzalez, D (1990) Ribosomal DNA restriction fragment length polymorphisms in *Rhizoctonia solani* *Phytopathology* 80 151–158
- Walden, R , Hayashi, H and Schell, J (1991) T-DNA as a gene tag *Plant J* 1 281–288
- Ward, E R and Jen, G C (1990) Isolation of single-copy-sequence clones from a yeast artificial chromosome library of randomly-sheared *Arabidopsis thaliana* DNA *Plant Mol Biol* 14 561–568
- Weber, J L (1990) Informativeness of human (dC-dA)_n-(dG-dT)_n polymorphisms *Genomics* 7 524–530
- Weber, J L and May, P E (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction *Am J Hum Genet* 44 388–396
- Welsh, J and McClelland, M (1990) Fingerprinting genomes using PCR with arbitrary primers *Nucl Acids Res* 18 7213–7218
- Welsh, J and McClelland, M (1991) Genomic fingerprints produced by PCR with consensus tRNA gene primers *Nucl Acids Res* 19 1–6
- Welsh, J , Honeycutt, R J , McClelland, M and Sobral, B W S (1991) Parentage determination in maize hybrids using the arbitrarily primed polymerase chain reaction (AP-PCR) *Theor Appl Genet* 82 473–476
- Williams, J G K , Kubelík, A R , Livak, K J , Rafalski, J A and Tingey, S V (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers *Nucl Acids Res* 18 6531–6535
- Williams, M N V , Pande, N , Nair, S , Mohan, M and Bennett, J (1991) Restriction fragment length polymorphism analysis of polymerase chain reaction products amplified from mapped loci of rice (*Oryza sativa* L.) genomic DNA *Theor Appl Genet* 82 489–498
- Wu, K K , Burnquist, W , Sorrells, M E , Tew, T L , Moore, P H and Tanksley, S D (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragments *Theor Appl Genet* 83 294–300
- Young, N D (1990) Potential applications of map-based cloning to plant pathology *Physiol Mol Plant Pathol* 37 81–94
- Young, N D , Zamir, D , Ganai, M W and Tanksley, S D (1988) Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the *Tm-2a* gene in tomato *Genetics* 120 579–585
- Yu, Z H , Mackill, D J , Bonman, J M and Tanksley, S D (1991) Tagging genes for blast resistance in rice via linkage to RFLP markers *Theor Appl Genet* 81 471–476
- Yuille, M A R , Goudie, D R , Affara, N A and Ferguson-Smith, M A (1991) Rapid determination of sequences flanking microsatellites *Nucl Acids Res* 19 1950–1951
- Zhao, X P and Kochert, G (1992) Characterization and genetic mapping of a short, highly repeated, interspersed DNA sequence from rice (*Oryza sativa* L.) *Mol Gen Genet* 231 353–359
- Zimmerman, P A , Langunnasch, N and Cullis, C A (1989) Polymorphic regions in plant genomes detected by an M13 probe *Genome* 32 824–828

3. Constructing a plant genetic linkage map with DNA markers

NEVIN DALE YOUNG

Department of Plant Pathology, 495 Borlaug Hall, University of Minnesota, St Paul, MN 55108, U S A

Contents

1 Overview	39	and physical maps	47
2 Constructing a linkage map with DNA markers	41	3 3 Completing a DNA genetic map	47
2 1 The mapping population	41	3 4 Parallel mapping in the same species	48
2 2 DNA polymorphisms among parents	42	3 5 Parallel mapping in related taxa	49
2 3 Choice of segregating population	43	4 Targeting specific genomic regions	49
2 4. Population size	44	5 Computer software for genetic mapping	51
2 5. DNA extractions	44	6 Perspectives on genetic mapping and DNA markers	53
3 Relationships among genetic maps	46	7 Acknowledgements	54
3 1 Relationship between DNA marker and cytogenetic maps	46	8 References	55
3 2 Relationship between genetic			

1. Overview

Scientists are constructing genetic linkage maps composed of DNA markers for many plant species today (Helentjaris 1987; Huen et al. 1991; McCouch et al. 1988; Tanksley et al. 1993). Two types of DNA markers have been widely used, restriction fragment length polymorphisms (RFLPs) (Botstein et al. 1980) (see Figs. 1 and 2) and random amplified polymorphic DNA markers (RAPDs) (Williams et al. 1990) (see Fig. 3). Both detect DNA polymorphisms and monitor the segregation of a DNA sequence among progeny of a genetic cross to construct a linkage map. While the theory of linkage mapping is the same for DNA markers as in classical genetic mapping, new considerations must be kept in mind. This is primarily a result of the fact that potentially unlimited numbers of DNA markers can be analyzed in a single mapping population. Backcross and F₂ populations are suitable for DNA-based mapping, but recombinant inbred (Burr and Burr 1991) and doubled haploid lines (Huen et al. 1991) provide permanent mapping resources. These types of populations may also be better suited for analysis of quantitative traits.

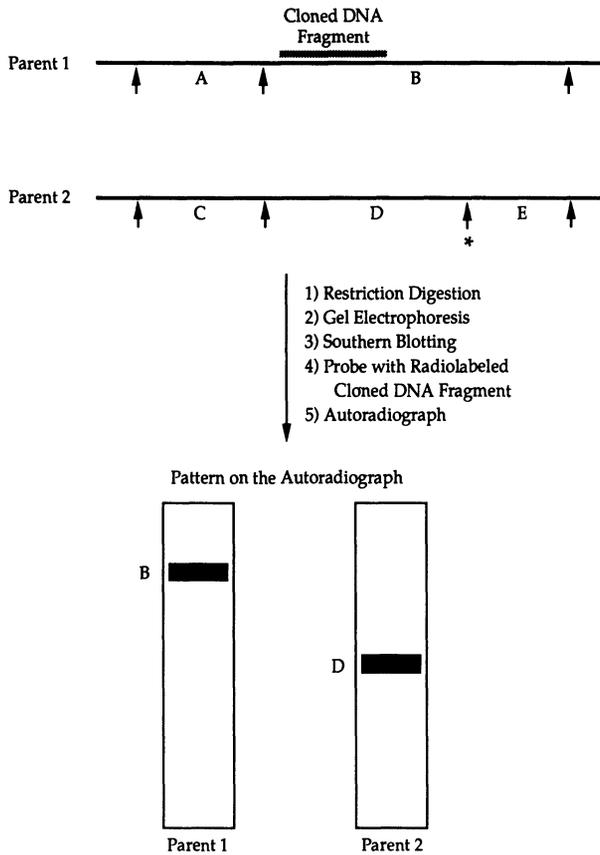


Fig. 1. Theory of restriction fragment length polymorphism (RFLP) analysis (from Young 1990). The solid, horizontal lines represent short segments of homologous DNA from two infertile parents. Arrows underneath the lines indicate restriction endonuclease recognition sites. Since the genomic sequences of both parents are nearly identical, most restriction sites are also identical. Occasionally, however, differences in DNA sequence lead to differences in restriction sites, denoted by a '*'. Restriction enzyme digestion of the DNA leads to fragments, labeled A through E. If the restriction digested DNA is size fractionated by gel electrophoresis, 'Southern' blotted, probed with the cloned DNA sequence, and autoradiographed, only fragments B and D are visible on the autoradiograph. These DNA fragments differ in size (due to the new restriction site '*' in parent B) and therefore migrate to different positions. The difference in migration between fragments B and D indicates a restriction fragment length polymorphism (RFLP). Fragments A, C, and E are not visible on the autoradiograph because they do not overlap the cloned DNA sequence.

The number of DNA markers on published linkage maps ranges all the way up to 1000 (Tanksley et al. 1993), so the resolution of DNA marker maps can be extremely high. However, alternate maps have been constructed in some species without links between maps and special effort is required to join the information (Beavis and Grant 1991). DNA-based maps can often be related to existing cytogenetic maps through the use of aneuploid or substitution lines (Helentjaris et al. 1986; Sharp et al. 1989; Young et al. 1987). Nonetheless, it remains difficult to know when one has actually mapped each chromosome with

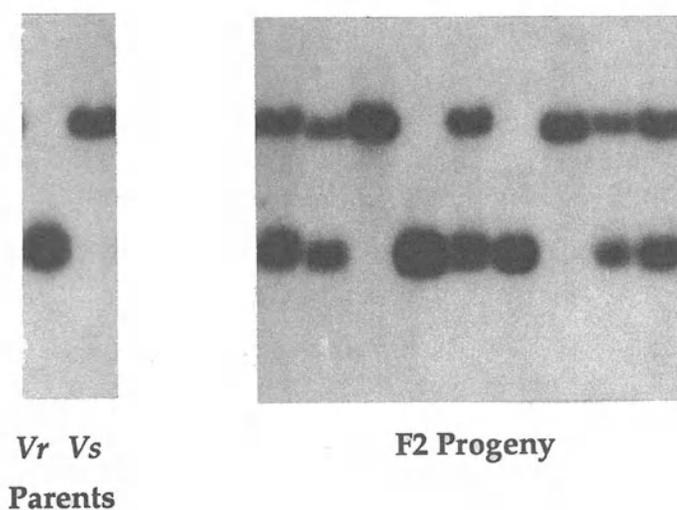


Fig 2 Segregation analysis of an RFLP marker in mungbean (*Vigna radiata*) Fifty-eight F₂ progeny from a cross between *Vigna radiata* (Vr) and *V. radiata* subsp. *sublobata* (Vs) were analyzed by RFLP analysis. The RFLP patterns for the parents and nine of the F₂ progeny are shown. Note that some F₂ progeny are homozygous for the RFLP pattern of Vr, others are homozygous for the pattern of Vs, and still others are heterozygous. This kind of information would typically be coded into a numeric form, entered into a computer, and compared to the segregation pattern data for hundreds of other RFLPs in order to construct a genetic linkage map of DNA markers.

DNA markers that go from one end to the other (Ganal et al. 1991). It is also uncertain how linkage maps of DNA markers compare to physical DNA maps. Recombination between DNA markers seems suppressed near centromeres, as well as heterochromatic and introgressed regions, leading to DNA genetic maps with markers clustered in some genomic regions and absent from others (Tanksley et al. 1993).

Applications of DNA markers to plant breeding and genetics have been described in previous reviews (Soller and Beckmann 1983; Tanksley et al. 1989). In this chapter, practical strategies for constructing genetic linkage maps using DNA markers will be described. Because of the breadth of this area, only an introduction to the concepts and techniques can realistically be covered.

2. Constructing a linkage map with DNA markers

2.1. The mapping population

The most critical decision in constructing a linkage map with DNA markers is the mapping population. In making this decision, several factors must be kept in mind, the most important of which is the goal of the mapping project. Is the goal simply to generate a framework map to provide a set of mapped loci for the

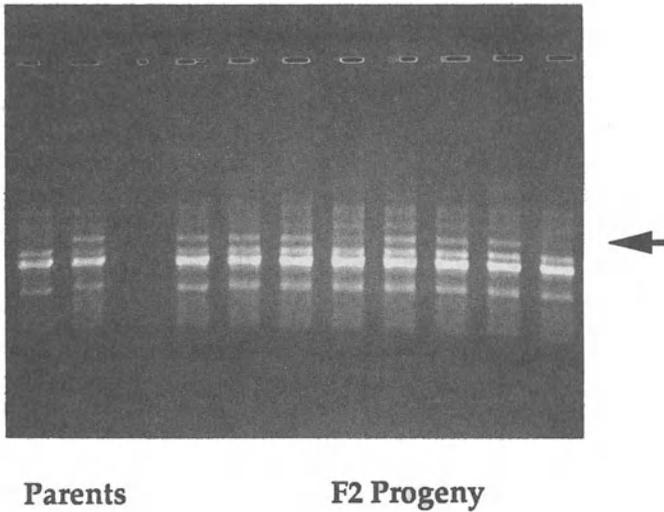


Fig. 3. Random Amplified Polymorphic DNA (RAPD) marker analysis. DNA samples from two tomato parents (left) and eight F_2 progeny (right) were analyzed by RAPD analysis (Williams et al. 1990) using a RAPD primer 10 nucleotides in length. Note that for each DNA sample, the RAPD reaction produced a few major bands, along with several minor bands. Among the major bands, one (arrow) differed between the parents and segregated among the F_2 progeny. This band is suitable for mapping as a DNA marker, as described in the text.

future, or instead, identify and orient DNA markers near a target gene for eventual map-based cloning? Perhaps the goal is mapping quantitative trait loci (QTL), or the monitoring of several disease resistance loci in the process of pyramiding them into a single background. Whichever goal is the motivating factor behind mapping, it will have a critical influence on which parents are chosen for crossing, the size of the population, how the cross is advanced, and which generations are used for DNA and phenotypic analysis.

2.2. *DNA polymorphisms among parents*

Sufficient DNA sequence polymorphisms between parents must be present. This cannot be overemphasized, for in the absence of DNA polymorphism, segregation analysis and linkage mapping are impossible. Naturally outcrossing species, such as maize, tend to have high levels of DNA polymorphisms and virtually any cross that does not involve related individuals will provide sufficient polymorphism for mapping (Helentjaris 1987). However in naturally inbreeding species, levels of DNA sequence variation are generally lower and finding suitable DNA polymorphisms may be more challenging (Miller and Tanksley 1990).

Sometimes mapping of inbreeding species requires that parents be as distantly related as possible. This can often be estimated based on geographical, morphological, or isozyme diversity. In some cases, suitable wide crosses may

already be available because a frequent goal in plant breeding in the past has been the introduction of desirable characters from wild relatives into cultivars. On the other hand, the requirement for sufficient DNA sequence polymorphism may preclude the use of DNA markers in some narrow-based crosses, such as between different cultivars in a crop species.

Extra effort and new technology may sometimes overcome the problems that result from a lack of DNA sequence polymorphism. In the case of RFLPs, the most obvious strategy is to test additional restriction enzymes. However, if polymorphisms are due to insertions and deletions, testing additional enzymes will not generally provide additional information. By contrast, if most RFLPs are due to base pair substitutions, testing additional enzymes may indeed be fruitful. In our lab, we have examined a cross between a cultivated tomato and an uncultivated, but closely related, red-fruited relative. Of 100 tomato genomic clones examined, 40 showed polymorphisms when tested with ten common restriction enzymes. Testing the remaining 60 clones with an additional 20 restriction enzymes uncovered 12 additional polymorphic clones (S. Aarons, unpub.).

New technologies may also provide better methods for uncovering polymorphisms within narrow-based crosses. For example, electrophoresis systems capable of separating DNA molecules with only a single base pair change have been described (Riedel et al. 1990). In some cases, probes based on minisatellites (Dallas 1988) or simple repeated tetranucleotide motifs (Weising et al. 1989) can uncover polymorphisms between closely related individuals. Because they are so variable at the DNA sequence level, these types of sequences may eventually provide markers useful for mapping in narrow-based crosses.

2.3. Choice of segregating population

Once suitable parents have been chosen, the type of genetic population to use for linkage mapping must be considered. Several different kinds of genetic populations are suitable. The simplest are F_2 populations derived from true F_1 hybrids and backcross populations. For most plant species, populations such as these are easy to construct, although sterility in the F_1 hybrid may limit some combinations of parents, particularly in wide crosses.

The major drawback to F_2 and backcross populations is that they are ephemeral, that is, seed derived from selfing these individuals will not breed true. This limitation can be overcome to a limited extent by cuttings, tissue culture or bulking F_3 plants to provide a constant supply of plant material for DNA isolation. Nevertheless, it is difficult or impossible to measure characters as part of QTL mapping in several locations or over several years with F_2 or backcross populations. For these reasons, permanent resources for genetic mapping are essential.

The best solution to this dilemma is the use of inbred populations that provide a permanent mapping resource. Recombinant inbred (RI) lines derived from individual F_2 plants are an excellent strategy (Burr et al. 1988; Burr and

Burr 1991). RI lines are created by single seed descent from sibling F_2 plants through at least five or six generations. This process leads to lines that each contain a different combination of linkage blocks from the original parents. The differing linkage blocks in each RI line provide a basis for linkage analysis. However, because several generations of breeding are required to generate a set of RIs, this process can be quite time-consuming. Moreover, some regions of the genome tend to stay heterozygous longer than expected from theory (Burr and Burr 1991) and obligate outcrossing species are much more difficult to map with RIs because of the inability to self plants.

Nevertheless, in cases where it is feasible, seed from RI lines is predominantly homogeneous and abundant, so the seed can be sent to any lab interested in adding markers to an existing linkage map previously constructed with the RI lines. Moreover, RI lines can be grown in replicated trials, several locations, and over several years – making them ideal for QTL mapping. Similar types of inbred populations, such as doubled haploids, can also be used for linkage mapping with many of the same advantages of RI lines (Huen et al. 1991).

2.4. *Population size*

Once an appropriate mapping population has been chosen, the appropriate population size must be determined. Since the resolution of a map and the ability to determine marker order is largely dependent on population size, this is a critical decision. Clearly, population size may be limited technically by how many seeds are available or by the number of DNA samples that can reasonably be prepared. Whenever possible, the larger the mapping population the better. Populations less than 50 individuals probably provide too little mapping resolution to be useful. Moreover, if the goal is high resolution mapping in specific genomic regions or mapping QTL of minor effect, much larger populations will be required. For example, Messeguer et al. (1991) examined over 1000 F_2 plants to construct a high resolution map around the *Mi* gene of tomato and Stuber et al. (1987) analyzed over 1800 maize F_2 s to find QTL controlling as little as 1% of the variation in yield components.

2.5. *DNA extractions*

No matter what type of population or DNA marker one plans to use, DNA must first be isolated from the plants in the mapping population. Fortunately, plants can be grown in a variety of environments and in different locations and still provide starting material for DNA isolation. This is in contrast to phenotypic markers, such as morphological or disease resistance traits, whose expression tend to be highly dependent upon growth conditions.

Several methods for DNA extraction suitable for DNA marker analysis have been developed (Dellaporta et al. 1983; Murray and Thompson 1984; Tai and Tanksley 1990). With each method, the goals have been simplicity, speed, and utilization of a small amount of starting material (Lamalay et al. 1990).

Simplicity and speed are absolutely essential for processing large numbers of individuals – an obvious necessity when large populations of several hundred, or even thousands, of individuals are to be examined. Small amounts of starting material are advantageous if larger quantities are hard to obtain, such as with seeds, seedlings, or physically small plants like *Arabidopsis*.

DNA used for genetic mapping does not need to be highly purified. As long as an extraction provides DNA in sufficient quantity and quality for restriction enzyme digestion or as a template for polymerase chain reaction (PCR), the method is probably satisfactory. Further efforts to purify DNA take time and cut down on the number of samples that can be processed. In general, limits to genetic mapping are more often due to small numbers of individuals in a mapping population than to DNA purity.

Still, one must guard against the most troublesome problems of RFLP and RAPD analysis. In the case of RFLPs, the major artifact is partial digestion of DNA. Since methods to extract DNA are streamlined, the DNA used in RFLP analysis can be quite impure. Sometimes this leads to partial digestion, which invariably leads to the appearance of extra bands upon autoradiography. It is very frustrating trying to map a 'polymorphic band' that turns out to be only a partial digest in one parent versus a complete digest in the other parent. However, complete digests of plant genomic DNA have a distinctive appearance upon gel electrophoresis, including a smear of DNA fragments throughout the appropriate size range for the restriction enzyme used, as well as the presence of reproducible DNA bands derived from the chloroplast. Moreover, partial digests lead to bands on autoradiographs that are generally fainter and higher in molecular weight than authentic restriction fragments.

Problems with RFLP mapping can also arise if too little DNA is used. Because RFLPs generally represent single copy sequences, the amount of any one target sequence in a genomic DNA sample can be vanishingly small. If too little DNA is loaded onto the gel for blotting, it may be impossible to see a signal after hybridization and autoradiography. Clearly, this will be related to genome size, and organisms with smaller genomes may require less DNA per sample than species with very large genomes. In practice, at least two micrograms, and potentially as much as ten micrograms or more of DNA should be used for RFLP analysis. Significantly, RAPDs require much less DNA as starting material – often an important advantage.

However, extra bands that do not represent real DNA polymorphisms can also be a problem in RAPD analysis. Because of the enormous amplification associated with PCR, as well as the tenuous association between RAPD primers and genomic DNA, variations in the set of amplified DNA molecules observed with a single primer are not uncommon. Variables as simple as differences in template and primer concentration can lead to the appearance or disappearance of DNA products. Moreover, PCR reactions in the absence of a plant DNA template can sometimes lead to products, possibly due to the synthesis of 'primer dimers' or even minute contamination of foreign DNA template. Because of these artifactual DNA products, special care must be taken to

optimize and standardize PCR reactions based on RAPDs. For these reasons, only the most prominent and dependable bands in a PCR reaction should typically be used for mapping.

3. Relationships among genetic maps

3.1. Relationship between DNA marker and cytogenetic maps

Once a DNA-based linkage map has been constructed, it should be correlated to chromosome karyotype or cytogenetic map, if possible. In the past, chromosomes in some plant species have been characterized and numbered based on their physical size and shape (Ramanna and Prakken 1967). In the scientific literature, communicating about specific chromosomes in these species is based on these chromosome designations (Rick 1971).

In these plant species, relating DNA marker maps to cytogenetic maps may be simple if phenotypic, resistance, or isozyme genes have already been assigned to specific chromosomes, through substitution or aneuploid analysis (Tang and Hart 1975; Hart 1979). In these situations, the linkage groups of a DNA marker map can be correlated to chromosomes simply by mapping the classical loci in terms of the DNA markers. However, in species where cytogenetic maps are unavailable, DNA marker maps will need to be related to specific chromosomes by different means. In the future, improved *in situ* hybridization may make it common to locate clones on specific chromosomes through direct observation.

For the present, the most common method to relate DNA marker maps to specific chromosomes is the use of aneuploids, such as monosomics (Helentjaris et al. 1986) and trisomics (Young et al. 1987), and substitution lines (Sharp et al. 1989). In species where aneuploid lines for each chromosome are available, nucleic acid hybridization with a mapped DNA clone indicates its chromosome location by observing the loss of a band (in the case of nullisomics) or a change in the relative signal on an autoradiogram (McCouch et al. 1988). This type of analysis may require 'within-lane' standards (such as a second DNA clone of previously determined chromosome location), so that subtle changes in the relative intensity of a band can be compared between lanes.

Using substitution lines to associate mapped DNA markers to specific chromosomes is similar in concept to aneuploid mapping. In cereal species where this approach is most common, lines with known chromosomes or chromosome arms substituted with homoeologous segments from alien species have been developed. Probing a DNA clone onto a blot containing restriction digested DNA from a complete set of substitution lines easily identifies the chromosome location of that clone (Sharp et al. 1989). This is because the substitution line corresponding to the location of a clone shows a different restriction fragment pattern compared to the other substitution lines.

3.2. *Relationship between genetic and physical maps*

Eventually, distances between DNA markers on a linkage map will be described not only by recombination frequency, but by actual physical distance. The need for this is obvious, for one of the primary goals of DNA-based genome maps is gene cloning based on chromosomal location, so-called 'map-based cloning' (Orkin 1986; Young 1990). However, there are currently very few published reports relating maps based on genetic recombination with physical distance in plants. In the future, this is certain to change for at least two reasons. First, the density of DNA marker linkage maps is constantly increasing. Simultaneously, techniques for analyzing very large DNA molecules, such as pulsed field gel electrophoresis (PFGE) (Schwartz and Cantor 1984) and clamped homogeneous electric field (CHEF) (Chu et al. 1986) electrophoresis, are becoming more powerful and widespread. Because the range of these techniques reaches three million base pairs or more (Chu et al. 1986), it is only a matter of time before large tracts of DNA linkage maps are correlated directly to physical maps. Related improvements in cloning vectors capable of carrying genomic inserts up to several hundred thousand base pairs (Burke et al. 1987) will also enhance attempts to relate genetic and physical maps.

In the few cases where information is available, the relationship between genetic and physical distance varies dramatically according to location on a chromosome (Ganal et al. 1989; Tanksley et al. 1992). In general, recombination is inhibited near centromeres and in heterochromatic and introgressed regions, while markers elsewhere appear to undergo relatively higher levels of recombination (Roberts 1965; Tanksley et al. 1992). This has the practical effect of making linkage maps appear to have many DNA markers clustered at a site probably corresponding to the centromere or heterochromatin, while in other parts of a map, markers are separated by large gaps, even after many hundreds of markers have been placed on the map. Despite the non-uniform distribution of markers in terms of recombination frequency, it is probable that the physical distance between markers is much more uniform.

3.3 *Completing a DNA genetic map*

If the relationship between a DNA linkage map and cytogenetically defined chromosomes can be determined, it should be easy to know when linkage groups corresponding to all chromosomes have been identified. However, for many plant species this will not be possible because cytogenetic maps are unavailable. In these cases, knowing when a map is 'complete' will be more challenging. One criterion should be that there are no remaining unlinked markers. Another criterion should be that the number of DNA marker linkage groups equals the number of chromosomes observed in the microscope, if known.

Even when this goal has been achieved, it is still uncertain whether the ends of chromosomes have been marked. However, the sequence and structure of

some plant chromosome telomeres have been determined (Ganal et al. 1991), and this may provide a method for mapping the ends of chromosomes. Moreover, satellite sequences tend to be located near the ends of chromosomes (Ganal et al. 1988). In the future, these sequences may provide a method for cloning adjacent low copy sequences, which could then be used to establish definitively the chromosome ends.

3.4. *Parallel mapping in the same species*

In the most important plant species there are often competing efforts to construct DNA-based genome maps. This has led to the unfortunate situation of having multiple maps for the same species with little or no information correlating one map to another. Of course this makes it difficult to relate the reported location of a gene on one map to its location on another map. It also means that the maps are less saturated, and therefore less powerful, than they could be.

Even where there is no proprietary barrier to relating maps to one another, there are often practical and theoretical problems. The most obvious is that markers polymorphic in one mapping population may not show variation in a second population. The first genetic maps were based on mapping populations optimized for DNA polymorphisms, often including parents from distinct, but cross-compatible species. As researchers move to more narrow crosses, previously excellent genetic markers will be useless for lack of polymorphism. When this happens it will be difficult to relate genetic map location between populations, except by cloning sequences that flank the original marker (a substantial amount of effort) or by testing adjacent DNA markers in hopes that they show more sequence variation.

A similar problem may be observed when one attempts to relate RAPD markers among different crosses. While there are often several bands observed in the analysis of each RAPD primer, only one of the bands may be polymorphic between two individuals (Williams et al. 1990). If an identical RAPD primer is analyzed in a second population, there is no guarantee that the same band (locus) will be the one that segregates. While any bands that do segregate in the second population will be suitable as markers, it is unlikely that they will map in the same region of the genome as the original marker. Similar situations can arise with RFLPs if they correspond to a sequence with multiple loci. Finally, there can be theoretical problems in relating linkage order data from one map to another, since each map is based on a different set of segregating individuals. However, the use of appropriate computer algorithms can potentially overcome this problem (S. Knapp, pers. comm.). Even with all these challenges, there have been successful attempts to relate maps from different crosses (Beavis and Grant 1991).

3.5. *Parallel mapping in related taxa*

One of the most powerful aspects of genetic mapping with DNA markers, particularly RFLPs, is the fact that markers mapped in one genus or species can often be used to construct parallel maps in related, but genetically incompatible, taxa. For this reason, a new mapping project can often build on previous mapping work in related organisms. Examples include a potato map constructed with tomato markers (Bonierbale et al. 1988; Gebhardt et al. 1991; Tanksley et al. 1993), a sorghum map constructed with maize markers (Hulbert et al. 1990), and a turnip map constructed with markers from cabbage (McGrath and Quiros 1991).

Not only does a pre-existing map provide a set of previously tested DNA markers, it also gives an indication of linkage groups and marker order. In the case of tomato and potato, only five paracentric inversions involving complete chromosome arms differentiate the two maps (Bonierbale et al. 1988; Gebhardt et al. 1991). Similar conservation of linkage order was observed between sorghum and maize (Hulbert et al. 1990). In cases like these, markers can be added to a new map in an optimum manner, either by focusing on markers evenly distributed throughout the genome, or by targeting specific regions of interest. In some cases, however, DNA clones may hybridize in multiple taxa, yet show little conservation in linkage group or order. Even though the tomato and potato maps are nearly homosequential (syntenic) in marker order, both differ significantly from the linkage map of pepper, despite the fact that all were constructed with the same RFLP markers (Tanksley et al. 1988).

4. **Targeting specific genomic regions**

In most cases, genome mapping is directed toward a comprehensive genetic map covering all chromosomes evenly. This is essential for effective marker-assisted breeding, QTL mapping, and chromosome characterization. However, there are special situations in which specific regions of the genome hold special interest. One example is where the primary goal of a research project is map-based cloning. In this case, markers that are very close to a target gene and suitable as starting points for chromosome walking are needed, so the goal is to generate a high density linkage map around that gene as quickly as possible. While the construction of a complete genome map by conventional means eventually leads to a high density map throughout the genome, special strategies for rapidly targeting specific regions have also been developed.

The first strategy for targeting specific regions was based on near isogenic lines (NILs). Over the years, breeders have utilized recurrent backcross selection to introduce traits of interest from wild relatives into cultivated lines. This process led to the development of pairs of NILs; one, the recurrent parent and the other, a new line resembling the recurrent parent throughout most of its genome except for the region surrounding the selected gene(s). This introgressed

region, derived from the donor parent and often highly polymorphic at the DNA sequence level, provides a target for rapidly identifying clones located near the gene of interest (Young et al. 1988; Martin et al. 1991; Paran et al. 1991; Muehlbauer et al. 1991). NILs make it easy to determine the location of a marker relative to the target gene. This is in contrast to typical genetic mapping where it would be necessary to test every clone with a complete mapping population to determine whether it mapped near the gene of interest.

Recently a strategy that makes it possible to target specific genomic regions without the need for developing specialized genotypes has been described (Michelmore et al. 1991; Giovanonni et al. 1992). The strategy is to select individuals from a segregating population that are homozygous for a trait of interest and pool their DNA. In the pooled DNA sample, the only genomic region that will be homozygous will be the region encompassing the gene of interest, which can then be used as a target for screening DNA markers rapidly. This means that any trait that can be scored in an F_2 , backcross, or RI population can now be rapidly targeted with DNA markers. Used in conjunction with RAPD markers, it is possible to identify large numbers of DNA markers in a region of interest in a short time.

Moreover, pooled DNA samples can also be generated based on homozygosity for a DNA marker (as opposed to a phenotypic trait). In this way, any genomic region of interest that has been previously mapped in terms of DNA markers can be rapidly targeted with new markers. This may be especially useful in trying to fill in gaps on a genetic map. All that is required is a pooled DNA sample selected on the basis of DNA markers flanking the genomic region of interest (Giovanonni et al. 1992).

Even as interesting regions become highly populated with DNA markers, there is a growing need for better methods to resolve marker order. Obviously, the easiest way to increase genetic resolution is to increase population size. Unfortunately, population sizes can quickly become unmanageable, particularly if resolution down to 0.1 centiMorgan (corresponding to one crossover every 1000 individuals) is required, as in map-based cloning. Again, pooling samples provides a solution.

Because the time-limiting factor in DNA marker analysis is the extraction of DNA from plant samples, plant samples from groups of five to ten individuals can be pooled and extracted in bulk, depending on the genome size of the species. The bulked DNA samples can then be tested for crossovers in the region of interest using flanking DNA markers. Those bulked samples showing crossovers are then tested with other DNA markers in the region of interest to resolve marker order. Plants in the bulked sample can also be examined individually to recover the individual with the crossover of interest (S. Tanksley, pers. comm.).

5. Computer software for genetic mapping

Advances in computer technology have been essential to progress in DNA marker maps. While the theory behind linkage mapping with DNA markers is identical to mapping with classical genetic markers, the complexity of the problem has increased dramatically. Linkage order is still based on maximum likelihood, in other words, the order of markers that yields the shortest distance and requires the fewest multiple crossovers between adjacent markers. Likewise, genetic distance between markers is measured in centiMorgans, which is based on the frequency of genetic crossing-over (accounting for the likelihood of, and interference among, multiple crossovers). For these reasons, the concepts and theories previously developed for classical genetic mapping can still be applied to mapping with DNA markers. The overwhelming difference between classical and DNA-based mapping lies in the number of markers that are mapped in a single population. With DNA-based genetic maps, this number can easily reach into the thousands, and so there is a close connection between progress in DNA markers and advances in computer technology.

In the simplest situations, all that is required to construct a linkage map from DNA marker data are statistics software packages capable of running Chi-squared contingency table analysis. This statistical test determines two-point linkage between markers, which can then form a basis for constructing linkage groups. Unfortunately, as the number of markers begins to grow, this approach becomes increasingly unsuited for comparing possible orders and choosing the best. Still, in research situations where computer power is limiting and where linkage analysis is based on relatively few markers, this strategy is perfectly suitable. The program, LINKAGE-1 (Suiter et al. 1983), carries out this type of genetic analysis automatically and also compares the observed allelic distributions to expected 3:1, 1:2:1, or 1:1:1:1 distributions. LINKAGE-1 requires a simple IBM microcomputer with little specialized upgrading. Other software packages that might be considered are SAS (SAS Institute Inc., Cary, NC) for the IBM and Statview (Abacus Concepts, Berkeley, CA) for the Macintosh.

For most mapping projects the most widely-used genetic mapping software is Mapmaker (Lander et al. 1987). Mapmaker is based on the concept of the LOD score, the 'log of the odds-ratio'. A LOD score indicates the log (10) of the ratio between the odds of one hypothesis (for example, linkage between two loci) versus an alternative hypothesis (no linkage in this example) (Morton 1955). Through the use of the LOD score, data from different populations can be pooled – one reason that the program has gained so much popularity among human and animal geneticists where population sizes can be limiting. Yet even in plants, Mapmaker has become a virtual standard for constructing genetic linkage maps, as indicated by its widespread use in the literature (Landry et al. 1987; Bonierbale et al. 1988; McCouch et al. 1988; Beavis and Grant 1991; Huen et al. 1991; Song et al. 1991).

Mapmaker's popularity for genetic analysis is based on the ease with which

it performs multipoint analysis of many linked loci. Most plant genetic linkage maps have at least one hundred markers, and eventually maps may have one thousand markers or more. Therefore, fast and simple multipoint analysis is absolutely essential to sort out the many different possible marker orders. Mapmaker has several routines that simplifies multipoint analysis, including an algorithm that quickly groups markers into likely linkage groups and another for guessing the best possible order. Once a plausible order has been established, another algorithm compares the strength of evidence for that order compared to possible alternatives in a routine called 'ripple'. The power of this routine is that it enables the user to confirm the best order in a way that increases only arithmetically with increasing number of loci (as opposed to factorially, if all possible orders must be compared). The main disadvantage of Mapmaker is that it requires a relatively powerful personal computer, such as a Unix-based computers like a Sun workstation or VAX minicomputer, or a Macintosh microcomputer with a math coprocessor and at least five megabytes of random access memory.

Just as new and exciting software is being developed for constructing genetic linkage maps, other types of software packages are being developed to make optimum use of the maps. One program, called 'Maize-RFLP Stack,' (D. Hoisington, pers. comm.), brings together all the information relating to an RFLP map of maize. Each DNA clone used for RFLP mapping is stored as a database record including essential information such as clone source, the insert size, vector, etc. The user can 'jump' directly from the database record to the position of the RFLP on a linkage map to see other markers in the vicinity and then find out more about the DNA clones that correspond to these nearby loci with the simple click of a computer mouse. Finally, a graphical image showing the banding pattern for each DNA clone on a standard set of maize inbred lines can then be displayed. This simplifies the task of determining which RFLPs are best to use in a cross and which restriction enzyme will be best for analysis.

In a very different strategy for optimizing the use of DNA marker information, the computer program, HyperGene, converts genotypic data into a 'graphical genotype' (Young and Tanksley 1989a,b). In a graphical genotype, the complete genome of an individual is displayed in a format that shows the parentage for each genomic interval graphically (see Fig. 4). This makes it possible to compare and contrast the genomic constitutions of different individuals with relative ease. Moreover, the user can use the computer mouse to 'paint' an ideal target genome. The program then examines the underlying DNA marker data and identifies individuals that have the selected genotype for the regions of interest, as well as which individuals are best for breeding toward that goal.

These computer programs demonstrate the close connection that is evolving between genetic analysis and computer technology. Recently, the U.S. Department of Agriculture established 'Genome Mapping Database Projects' for four important crop species, maize, wheat, soybean, and pine. These databases will incorporate many of the routines mentioned above, as well as

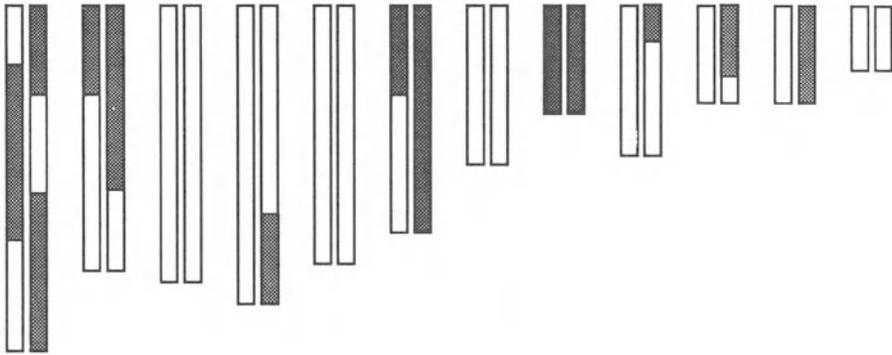


Fig. 4. Graphical genotype for a hypothetical individual from an F_2 population. This type of population can be developed in only two generations and has segregating 'linkage blocks' (shown in dark) that provide a basis for genetic mapping. Unfortunately, F_2 individuals are ephemeral, which means selfed seed derived from these individuals will be a different genotype.

many others that are being developed. The databases will also attempt to relate different DNA marker maps constructed within a single species to one another. The databases are being established in a coordinated manner, so that progress in one of the systems can be quickly shared with the others. Because this will be a national database, researchers throughout the country will be able to access the information simply by connecting to national communication networks that have recently been established. Since the databases will be graphically-based and use 'click-and-point' routines, they should be easy to learn and use. Eventually, these databases will provide a foundation for the establishment of genome databases in other important crop and plant research systems.

6. Perspectives on genetic mapping and DNA markers

In the brief period since DNA marker technology was first applied to plants, there has been an explosion in the development and application of genetic linkage maps. Using these new DNA-based maps, researchers have constructed maps in species where only poorly populated classical maps existed before (Bonierbale et al. 1988; Gebhardt et al. 1991; Landry et al. 1987), located genes for qualitative (Young et al. 1988; Landau-Ellis et al. 1991) and quantitative (Paterson et al. 1988; Keim et al. 1990) characters, often in great detail (Messeguer et al. 1991; Paterson et al. 1990), and taken the first steps toward gene cloning based solely on genetic map position (Ganal et al. 1989). Despite this incredible progress, DNA marker technology has a very long way to go before its full potential is realized.

With current procedures, the number of plant samples and DNA markers that can reasonably be processed limits the widespread application of mapping technology. Even the most efficient DNA extraction techniques handle only 100

or at most 200 samples each day, and once samples have been isolated, significant investments in time and effort are still required to obtain genotypic information. Given the fact that a typical breeding project might include several thousand, or even tens of thousands of individuals, and since information is needed as quickly as possible to make breeding decisions, the current technical limitations are significant. These limitations also constrain the application of DNA marker technology in QTL mapping to genetic factors with relatively major effects, and in map-based cloning to plant species that have small genomes. Finally, DNA marker technology is still so technically complex that it is practically impossible for it to be applied where it is needed most – in less-developed countries.

However, better types of DNA genetic markers are on the horizon. The recent development of RAPD markers gives an indication of the future. Even though DNA samples must still be prepared for each individual, processing each sample is technically simple and amenable to automation. At the same time, advances, such as gel electrophoresis systems capable of distinguishing single base pair changes (Riedel et al. 1990), are likely to expand the concept of DNA markers significantly. DNA sequencing technology itself is progressing so rapidly that it may soon be practical to sequence DNA segments directly as part of segregation analysis and genetic mapping studies.

Even as DNA marker technology advances, parallel achievements are essential in complementary technologies. As the number of markers, genetic resolution, and amount of mapping information grows, so does the need for better computer algorithms and databases. For physical mapping and map-based cloning to succeed, techniques for analyzing, manipulating, and cloning very large fragments of DNA must be improved significantly. Finally, genetic linkage maps, even those based on DNA markers, are still limited by the range of sexual crosses that can be made. To make the most of genes uncovered through genetic mapping, improvements in making wide crosses, somatic hybrids, and plant transformation will be essential. In the future, better DNA markers, along with advances in these complementary technologies, will enable linkage mapping, one of the oldest genetic techniques, to become one of the most powerful.

7. Acknowledgements

I wish to thank Dr. Deborah Samac and Dr. Desiree Menancio-Hautea for reviewing the manuscript. This work was supported, in part, by grants from the Rockefeller Foundation, the U.S.D.A. Plant Genome Mapping Database project, the U.S.-A.I.D. Special Constraints program, and the Minnesota Soybean Promotion and Research Council. This paper is published as contribution No. 20,249 of the series of the Minnesota Agricultural Experiment Station on research conducted under Project 015, supported by G.A.R. funds.

8. References

- Beavis, W D and Grant, D (1991) A linkage map based on information from four F₂ populations of maize (*Zea mays* L) *Theor Appl Genet* 82 636–644
- Bonierbale, M W, Plaisted, R L and Tanksley, S D (1988) RFLP maps based on common sets of clones reveal modes of chromosome evolution in potato and tomato *Genetics* 120 1095–1103
- Botstein, D, White, R L, Skolnick, M and Davis, R W (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms *Am J Hum Genet* 32 314–331
- Burke, D T, Carle, G F and Olson, M V (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors *Science* 236 806–812
- Burr, B and Burr, F A (1991) Recombinant inbreds for molecular mapping in maize *Trends Genet* 7 55–60
- Burr, B, Burr, F A, Thompson, K H, Albertson, M C and Stuber, C W (1988) Gene mapping with recombinant inbreds in maize *Genetics* 118 519–526
- Chu, G, Volrath, D and Davis, R W (1986) Separation of large DNA molecules by contour-clamped homogenous electric fields *Science* 234 1582–1585
- Dallas, J F (1988) Detection of DNA ‘fingerprints’ of cultivated rice by hybridization with a human minisatellite DNA probe *Proc Natl Acad Sci U S A* 85 6831–6835
- Dellaporta, S L, Wood, J and Hicks, J B (1983) A plant DNA minipreparation version II *Plant Mol Biol Rept* 1 19–21
- Ganal, M W, Lapitan, N L V and Tanksley, S D (1988) A molecular and cytogenetic survey of major repeated DNA sequences in tomato (*Lycopersicon esculentum*) *Mol Gen Genet* 213 262–268
- Ganal, M W, Young, N D and Tanksley, S D (1989) Pulsed field gel electrophoresis and physical mapping of large DNA fragments in the *Tm-2a* region of chromosome 9 in tomato *Mol Gen Genet* 215 395–400
- Ganal, M W, Lapitan, N L V and Tanksley, S D (1991) Macrostructure of the tomato telomeres *Plant Cell* 3 87–94
- Gebhardt, C, Ritter, E, Barone, T, Debener, T, Walkemeier, B, Schachtschabel, U, Kaufman, H, Thompson, R D, Bonierbale, M W, Ganal, M W, Tanksley, S D and Salamini, F (1991) RFLP maps of potato and their alignment with the homoeologous tomato genome *Theor Appl Genet* 83 49–57
- Giovanonni, J, Wing, R and Tanksley, S D (1992) Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations *Nucl Acids Res* 19 6553–6558
- Hart, G E (1979) Genetical and chromosomal relationships among the wheat and their relatives *Stadler Genet Symp* 11 9–11
- Helentjaris, T (1987) A genetic linkage map for maize based on RFLPs *Trends Genet* 3 217–221
- Helentjaris, T, Weber, D F and Wright, S (1986) Use of monosomics to map cloned DNA fragments in maize *Proc Natl Acad Sci U S A* 83 6035–6039
- Huen, M, Kennedy, A E, Anderson, J A, Lapitan, N L V, Sorrells, M E and Tanksley, S D (1991) Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*) *Genome* 34 437–447
- Hulbert, S H, Richter, T E, Axtell, J D and Bennetzen, J L (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes *Proc Natl Acad Sci U S A* 87 4251–4255
- Keim, P, Diers, B W, Olson, T C and Shoemaker, R C (1990) RFLP mapping in soybean association between marker loci and variation in quantitative traits *Genetics* 126 735–742
- Lamalay, J C, Tejwani, R and Rufener, G K I (1990) Isolation and analysis of genomic DNA from single seeds *Crop Sci* 30 1079–1084
- Landau-Ellis, D, Angermuller S, Shoemaker, R and Gresshoff, P M (1991) The genetic locus controlling supernodulation in soybean (*Glycine max* L) co-segregates tightly with a cloned molecular marker *Mol Gen Genet* 228 221–226
- Lander, E S, Green, P, Abrahamson, J, Barlow, A, Daly, M J, Lincoln, S E and Newburg, L

- (1987) MAPMAKER, an interactive computer program for constructing genetic linkage maps of experimental and natural populations *Genomics* 1 174
- Landry, B S , Kesseli, R V , Farrara, B and Michelmore, R W (1987) A genetic map of lettuce (*Lactuca sativa* L) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers *Genetics* 116 331-337
- Martin, G B , Williams, J G K and Tanksley, S D (1991) Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato using random primers and near-isogenic lines *Proc Natl Acad Sci U S A* 88 2336-2340
- McCouch, S R , Kochert, G , Yu, Z Y , Khush, G S , Coffman, W R and Tanksley, S D (1988) Molecular mapping of rice chromosomes *Theor Appl Genet* 76 815-829
- McGrath, J M and Quiros, C F (1991) Inheritance of isozyme and RFLP markers in *Brassica campestris* and comparison with *B oleracea* *Theor Appl Genet* 82 668-673
- Messeguer, R , Ganal, M , de Vicente, M C , Young, N D , Bolkan, H and Tanksley, S D (1991) High resolution RFLP map around the root knot nematode resistance gene (*Mi*) in tomato *Theor Appl Genet* 82 529-536
- Michelmore, R W , Paran, I and Kesseli, R V (1991) Identification of markers linked to disease-resistance genes by bulked segregation analysis a rapid method to detect markers in specific genome regions by using segregating populations *Proc Natl Acad Sci U S A* 88 9828-9832
- Miller, J C and Tanksley, S D (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon* *Theor Appl Genet* 80 437-448
- Morton, N E (1955) Sequential tests for the detection of linkage *Am J Hum Genet* 7 277-318
- Muehlbauer, G J , Staswick, P E , Specht, J E , Graef, G L , Shoemaker, R C and Keim, P (1991) RFLP mapping using near-isogenic lines in the soybean [*Glycine max* (L) Merr] *Theor Appl Genet* 81 189-198
- Murray, M and Thompson, W F (1984) Rapid isolation of high molecular weight plant DNA *Nucl Acids Res* 8 4321-4325
- Orkin, S H (1986) Reverse genetics and human disease *Cell* 47 845-850
- Paran, I , Kesseli, R and Michelmore, R (1991) Identification of restriction fragment length polymorphism and random amplified polymorphic DNA markers linked to downy mildew resistance genes in tomato, using near-isogenic lines *Genome* 34 1021-1027
- Paterson, A H , DeVerna, J W , Lanin, B and Tanksley, S D (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes in an interspecies cross of tomato *Genetics* 124 735-742
- Paterson, A H , Lander, E S , Hewitt, J D , Peterson, S , Lincoln, S E and Tanksley, S D (1988) Resolution of quantitative traits into Mendelian factors using a complete linkage map of restriction fragment length polymorphisms *Nature (London)* 335 721-726
- Ramanna, M S and Prakken, R (1967) Structure and homology between pachytene and somatic chromosomes of the tomato *Genetics* 38 115-133
- Rick, C M (1971) Some cytogenetic features of the genome in diploid plant species *Stadler Genet Symp* 1 153-174
- Riedel, G E , Swanberg, S L , Kuranda, K D , Marquette, K , LaPan, P , Bledoe, P , Kennedy, A and Lin, B Y (1990) Denaturing gradient gel electrophoresis identifies genomic DNA polymorphism with high frequency in maize *Theor Appl Genet* 80 1-10
- Roberts, P A (1965) Difference in the behavior of eu- and hetero-chromatin crossing over *Nature (London)* 205 725-726
- Rommens, J M , Lannuzzi, M C , Kerem, B , Drumm, M L , Melmer, G , Dean, M , Rozmahel, R , Cole, J L , Kennedy, D , Hidaka, N , Zsiga, M , Buchwald, M , Riordan, J R , Tsui, L and Collins, F S (1989) Identification of the cystic fibrosis gene chromosome walking and jumping *Science* 245 1059-1065
- Schwartz, D C and Cantor, C R (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis *Cell* 37 67-75
- Sharp, P J , Chao, S , Desai, S and Gale, M D (1989) The isolation, characterization and application in the Triticeae of a set of wheat RFLP probes identifying each homoeologous chromosome arm *Theor Appl Genet* 78 342-348

- Soller, M and Beckmann, J S (1983) Genetic polymorphism in varietal identification and genetic improvement *Theor Appl Genet* 67 837-843
- Song, K M , Suzuki, J Y , Slocum, M K , Williams, P H and Osborn, T C (1991) A linkage map of *Brassica rapa* (syn *campestris*) based on restriction fragment length polymorphism loci *Theor Appl Genet* 82 296-304
- Stuber, C W , Edwards, M D and Wendel, J F (1987) Molecular marker-facilitated investigations of quantitative trait loci in maize II Factors influencing yield and its component traits *Crop Sci* 27 639-648
- Sutler, K A , Wendel, J F and Case, J S (1983) Linkage-1 A PASCAL computer program for the detection and analysis of genetic linkage *J Hered* 74 203-204
- Tai, T H and Tanksley, S D (1990) A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue *Plant Mol Biol Rep* 8 297-303
- Tang, K S and Hart, G E (1975) Use of isozymes as chromosome markers in wheat-rye addition lines and triticales *Genet Res* 26 187-201
- Tanksley, S D , Bernatzky, R , Lapitan, N L and Prince, J P (1988) Conservation of gene repertoire but not gene order in pepper and tomato *Proc Natl Acad Sci U S A* 85 6419-6423
- Tanksley, S D , Young, N D , Paterson, A H and Bonierbale, M W (1989) RFLP mapping in plant breeding new tools for an old science *Bio/Technology* 7 257-264
- Tanksley, S D , Ganal, M W , Prince, J P , deVincente, M C , Bonierbale, M W , Broun, P , Fulton, T M , Giovannoni, J J , Grandillo, S , Martin, G B , Messeguer, R , Miller, J C , Miller, L , Paterson, A H , Pineda, O , Roder, M , Wing, R A , Wu, W and Young, N D (1992) High density molecular linkage maps of the tomato and potato genomes biological inferences and practical applications *Genetics* 132 1141-1160
- Weising, K , Weigand, F , Driesel, A , Kahl, G , Zischler H and Epplen, J T (1989) Polymorphic GATA/GACA repeats in plant genomes *Nucl Acids Res* 17 10128
- Williams, J , Kubelik, A , Livak, K , Rafalski, J and Tingey, S (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers *Nucl Acids Res* 18 6531-6535
- Young, N D (1990) Potential applications of map-based cloning to plant pathology *Physiol Mol Plant Pathol* 37 81-94
- Young, N D and Tanksley, S D (1989a) Graphics-based whole genome selection using RFLPs In T Helentjaris and B Burr (eds), *Current Communications in Molecular Biology Development and Application of Molecular Markers to Problems in Plant Genetics*, pp 123-129 Cold Spring Harbor Press, Cold Spring Harbor, NY
- Young, N D and Tanksley, S D (1989b) Restriction fragment length polymorphism maps and the concept of graphical genotypes *Theor Appl Genet* 77 95-101
- Young, N D , Miller, J and Tanksley, S D (1987) Rapid chromosomal assignment of multiple genomic clones in tomato using primary trisomics *Nucl Acids Res* 15 9339-9348
- Young, N D , Zamir, D , Ganal, M W and Tanksley, S D (1988) Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the *Tm-2a* gene in tomato *Genetics* 120 579-585

4. Mapping quantitative trait loci

STEVEN J. KNAPP

Department of Crop and Soil Science and Center for Gene Research and Biotechnology, Oregon State University, Corvallis, OR 97331, U S A

Contents

1. Introduction	58	6. Selecting sources of favorable alleles and marker-assisted selection	85
2. Experiment design	60	7. QTL genotype by environment interactions	90
3. Multiple environment experiments	72	8. Acknowledgements	93
4. Replications of lines and replications of QTL genotypes	82	9. References	93
5. Multiple loci	84		

1. Introduction

Different alleles at *quantitative trait loci* (QTL) cause genetic differences between individuals and families for quantitative traits (Bulmer 1980; Falconer 1981). QTL genotypes cannot be determined by inspecting the distributions of trait phenotypes alone. This is one of the fundamental problems of quantitative genetics. Historically important quantitative genetic parameters, e.g., additive genetic variance and heritability, summarize differences between alleles at QTL, but do not shed light on the genetics of QTL. Methods for mapping QTL are needed to achieve this. QTL are mapped by using genetic markers linked to QTL to draw inferences about differences between alleles at QTL.

Several parameter estimation methods, mostly variants of two classes of *interval mapping* methods, have been described for mapping QTL using matings between inbred lines (Carbonell et al. 1992; Haley and Knott 1992; Jansen 1992; Jensen 1989; Knapp et al. 1990; Knapp et al. 1991; Knott and Haley 1992; Lander and Botstein 1989; Lou and Kearsy 1989; Martinez and Curnow 1993; Simpson 1989; Van Ooijen 1992; Weller 1986). While interval mapping should not be narrowly defined, it has most often meant using marker-brackets or flanking markers as independent variables to build a genetic model for testing the hypothesis of no QTL against the hypothesis of one QTL within a marker-bracket (differentiating between two genetic models). A great many other

genetic models can be hypothesized and tested, e.g., see Knapp (1991) and Martinez and Curnow (1993), and these use the same principles.

Least squares (Haley and Knott 1992; Knapp et al. 1990) and maximum likelihood (Carbonell et al. 1992; Jensen 1989; Lander and Botstein 1989; Lou and Kearsey 1989; Simpson 1989) methods are the two chief classes of interval mapping methods. The fundamentals of maximum likelihood methods for interval mapping have been extensively reviewed (Lander and Botstein 1989; Knott and Haley 1992; Van Ooijen 1992; Weller 1993). As Haley and Knott (1992) show, the power of least squares and ML methods is equal for matings between inbred lines.

ML interval mapping methods have been criticized for failing to find QTL when multiple QTL are segregating (Martinez and Curnow 1993). Martinez and Curnow (1993), for example, show that interval mapping (testing the hypothesis of one QTL versus no QTL) can fail when two linked QTL are segregating. This is not surprising since what is needed is a test of the hypothesis of two linked QTL versus one QTL or no QTL. The problem has nothing to do with the parameter estimation method *per se*. Rather it has to do with the inadequacy of the underlying genetic model, or the failure to test hypotheses about different genetic models.

The distinction between the genetic model and the parameter estimation method has to be kept clear. Every parameter estimation method is prone to failure when the stated genetic model is wrong. The challenge is to model the genetics adequately and to develop statistics for differentiating between genetic models. Unless software is developed to do this, most investigators are not likely to test hypotheses about different multilocus genetic models.

Least squares methods are more suited to handling a wide range of parameter estimation problems than ML methods, such as estimating the parameters of multiple QTL, efficiently searching genomes for multiple QTL, and estimating QTL parameters from multiple environment experiments done using replicated experiment designs, e.g. randomized complete blocks. At least the computations are more straightforward. The problem of handling different experiment and environment designs has not been thoroughly addressed for any QTL mapping method (Jansen 1992; Knapp and Bridges 1990; Soller and Beckmann 1990). And the problem of how to most effectively execute genome searches is not completely settled either, although the fundamentals are clear (Knapp et al. 1993).

Genetic models for mapping QTL are functions of frequencies of observed marker phenotypes and hypothesized QTL phenotypes. They are defined by the pedigree of the individuals, lines, or families, the number of quantitative trait loci, whether or not any of the QTL are linked, and the genetic map of the population. Several genetic models are routinely tested when mapping QTL, although the ultimate objective of any experiment is to find the most satisfactory multilocus genetic model for a given trait. The joint frequencies of marker and hypothesized QTL genotypes must be defined to estimate the parameters of multilocus models. This becomes very cumbersome as the number of loci

increases. This problem is closely tied to the problem of how to search a genome and how to get unbiased test statistics (Knapp et al. 1993).

An important group of mating designs for plant breeders uses inbred lines as parents. Some of the progeny types which can be developed from matings between inbred lines are F_2 , F_3 , $F_{2,4}$, $F_{3,4}$, $F_{2,5}$, $F_{3,5}$, $F_{4,5}$, backcross (BC), BC_1S_1 , doubled haploid (DH), recombinant inbred (RI), and an assortment of testcross progeny. Genetic models and methods for mapping QTL have been defined for some of these progeny types, but many gaps remain. Defining these genetic models is nonetheless straightforward for matings between inbred lines.

2. Experiment design

Most of the QTL mapping methods and software developed thus far can be used for experiments where unreplicated progeny are tested or for experiments where replicated progeny are tested and completely randomized (CR) experiment designs are used. MAPMAKER-QTL (Lander and Botstein 1989), for example, cannot be used to estimate QTL parameters from a randomized complete blocks experiment design or from multiple environment experiments without ignoring blocks or environments. Nor can linear regression interval mapping methods (Haley and Knott 1992; Curnow and Martinez 1992) be used for these experiments, although they can be extended without any difficulty. There are two separate problems to address. One is developing methods to handle a range of experiment, environment, and mating designs. This is what we address by using linear least squares interval mapping (LIM). The other is developing software to implement these methods.

The acronym LIM is used for the entire group of linear least squares interval mapping methods whether parameters are estimated using linear regression *per se* (Haley and Knott 1992; Curnow and Martinez 1992), regression on dummy variables with QTL genotype coefficients used as covariables, or regression on dummy variables with QTL genotype coefficients used to define linear contrasts among marker genotype means, as is done throughout this paper.

Several QTL mapping advances have been described where parameter estimation was done using standard statistics software, e.g., GENSTAT or SAS; however, none of it has been automated (Knapp 1989; Knapp et al. 1993; Jensen 1991). The greatest needs for plant breeders are methods and software for estimating QTL and other genetic parameters for balanced and unbalanced linear models for different experiment, mating, and environment designs. LIM is undoubtedly the most effective way to handle these estimation problems.

LIM methods are developed below for virtually any experiment or environment design by bringing together linear least squares (Haley and Knott 1992) and linear model theory for unbalanced linear models (Knapp et al. 1993). The example used is the randomized complete blocks experiment design. Suppose an experiment is done where lines of some sort, e.g., doubled haploid

Table 1. Degrees of freedom, Type III sum of squares, and expected mean squares for lines tested in randomized complete blocks in one environment where the effects of QTL genotypes (QTL) are fixed and the effects of lines and blocks are random and factors other than QTL genotypes are balanced.

Factor	Degrees of freedom ^a	Sum of squares	Expected mean square ^b
Block	$df_R = b - 1$	$R[b \mu, g]$	$E(M_B) = \sigma_E^2 + N\sigma_B^2$
Line (G)	$df_G = N - 1$	$R[g \mu, b]$	$E(M_G) = \sigma_E^2 + r\sigma_G^2$
QTL (Q)	$df_Q = q - 1$	$R[q \mu, b]$	$E(M_Q) = \sigma_E^2 + r\sigma_G^2 + \phi_Q^2$
G:Q	$df_{G:Q} = N - q$	$R[g(q) \mu, g, b]$	$E(M_{G:Q}) = \sigma_E^2 + r\sigma_G^2$
Residual	$df_E = (N-1)(b-1)$	$R[e \mu, b, g]$	$E(M_E) = \sigma_E^2$

^a q is the number of QTL genotypes, $N = \sum_{i=1}^q n_i$ is the number of lines where n_i is the number of lines of the i th QTL genotype, and b is the number of blocks.

^b σ_E^2 is the error variance, σ_G^2 is the line nested in QTL genotype variance, σ_G^2 is the between line variance, ϕ_Q^2 is the variance of fixed effects of QTL genotypes, and

$$\bar{n} = \frac{N - \frac{\sum n_i^2}{N}}{q - 1}$$

or F_3 lines, are tested in randomized complete blocks in one environment (Tables 1). The linear model

$$y_{ij} = \mu + b_i + g_j + e_{ij}$$

can be used to estimate the usual quantitative genetic parameters, e.g., line means and between line variances, where y_{ij} is the ij th observation of the quantitative trait, μ is the population mean, b_i is the effect of the i th block, g_j is the effect of the j th line and e_{ij} is the random error for the j th line in the i th block. The effects of lines and blocks are random (Table 1). The objective of such an experiment is usually to select the most outstanding lines. Other objectives might be to test the null hypothesis of no between line variance ($H_0: \sigma_G^2 = 0$) and to estimate heritabilities and expected selection gains. Methods for estimating the parameters of (1) are straightforward.

Suppose, for example, doubled haploid lines are tested. The genetic variance between doubled haploid lines is

$$\sigma_G^2 = 2\sigma_A^2 = [E(M_G) - E(M_E)]/r,$$

while the line-mean heritability for selection among doubled haploid lines is

$$H = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2/r} = \frac{2\sigma_A^2}{2\sigma_A^2 + \sigma_E^2/r} = 1 - \frac{E(M_E)}{E(M_G)}$$

where σ_A^2 is the additive genetic variance and epistasis is ignored (Table 1). Without a genetic map and marker phenotypes for several loci dispersed throughout the genome, this is where the analysis of this experiment might end.

The addition of a genetic map and molecular marker phenotypes for several loci, however, creates the basis for estimating the parameters of genes or QTL underlying differences between lines, for which additional genetic models and estimation methods are needed.

Classical quantitative genetic parameters can be defined as functions of marker or QTL parameters by using linear models. The effects of lines, for example, can be defined as a function of the effects of their QTL (Knapp and Bridges 1990). Rewriting (1) to show this gives

$$y_{ikj} = \mu + b_i + q_k + g(q)_{kj} + e_{ikj}$$

where y_{ikj} is the ikj th observation of the quantitative trait, μ is the population mean, b_i is the effect of the i th block, q_k is the effect of the k th QTL genotype, $g(q)_{kj}$ is the effect of the j th line nested in the k th QTL genotype, and e_{ikj} is the random error for the j th line of the k th QTL genotype in the i th block. The effects of QTL genotypes are fixed, while the effects of other factors are random.

The effects of lines are the sum of the effects of QTL genotypes and lines nested in QTL genotypes ($g_j = q_k + g(q)_{kj}$). The latter are the effects between lines which are left over after estimating the effects between QTL genotypes. They are the effects of all of the genes which are not part of the model. This is easily seen by examining the expected mean squares. The sum of squares between lines (S_G) from (1) is the sum of sum of squares between QTL genotypes (S_Q) and between lines nested in QTL genotypes (S_{GQ}) from (2) (Knapp and Bridges 1990); thus,

$$\begin{aligned} E(M_G) &= \sigma_e^2 + r\sigma_G^2 = \frac{df_Q E(M_Q) + df_{GQ} E(M_{GQ})}{df_Q + df_{GQ}} \\ &= \sigma_E^2 + r\sigma_{GQ}^2 + \frac{r\bar{n}(q-1)}{N-1} \sigma_Q^2 \end{aligned} \quad (3)$$

where σ_Q^2 is the genetic variance between QTL genotypes.

The QTL genotypes of (2) cannot be observed so the frequencies of hypothesized QTL genotypes must be inferred from the marker phenotypes to estimate QTL genotype effects (q_i). This is the essential feature of QTL mapping. A specific genetic model must be used to illustrate how this is done using least squares, but the principles work for virtually any genetic model. To review the theory and illustrate LIM, a genetic model for doubled haploid lines is used where a QTL is hypothesized to lie between two linked marker loci A and B and there is no interference, i.e., the coefficient of coincidence (γ) is equal to 1.0 (Table 2). Genetic models for the doubled haploid mating design are well known (Knapp et al. 1990; Knapp 1991).

Genetic models for QTL mapping problems are nonlinear; however, they can be made linear by fixing the recombination frequencies between the QTL and the markers (θ_{AQ} and θ_{BQ}) (Haley and Knott 1992; Knapp et al. 1990; Lander and Botstein 1989). Under null interference ($\gamma = 1$),

Table 2. Expected marker genotypes means (μ_m) for doubled haploid lines under no interference where A and B are codominant marker loci with alleles A and a and B and b, respectively, Q is a hypothesized quantitative trait locus with alleles Q and q, the locus order is AQB, θ_{AB} is the recombination frequency between A and B, θ_{AQ} is the recombination frequency between A and Q, θ_{BQ} is the recombination frequency between B and Q, μ_{QQ} is the mean of QQ genotypes at the QTL, μ_{qq} is the mean of qq genotypes at the QTL, and m indexes marker genotypes.

Marker genotype		Expected marker genotype mean
A	B	
AA	BB	$\mu_1 = \frac{(1 - \theta_{AQ} - \theta_{BQ} + \theta_{AQ}\theta_{BQ})\mu_{QQ} + \theta_{AQ}\theta_{BQ}\mu_{qq}}{1 - \theta_{AQ} - \theta_{BQ} + 2\theta_{AQ}\theta_{BQ}}$
AA	bb	$\mu_2 = \frac{(\theta_{BQ} - \theta_{AQ}\theta_{BQ})\mu_{QQ} + (\theta_{AQ} - \theta_{AQ}\theta_{BQ})\mu_{qq}}{\theta_{AQ} + \theta_{BQ} - 2\theta_{AQ}\theta_{BQ}}$
aa	BB	$\mu_3 = \frac{(\theta_{AQ} - \theta_{AQ}\theta_{BQ})\mu_{QQ} + (\theta_{BQ} - \theta_{AQ}\theta_{BQ})\mu_{qq}}{\theta_{AQ} + \theta_{BQ} - 2\theta_{AQ}\theta_{BQ}}$
aa	bb	$\mu_4 = \frac{\theta_{AQ}\theta_{BQ}\mu_{QQ} + (1 - \theta_{AQ} - \theta_{BQ} + \theta_{AQ}\theta_{BQ})\mu_{qq}}{1 - \theta_{AQ} - \theta_{BQ} + 2\theta_{AQ}\theta_{BQ}}$

$$\theta_{AB} = \theta_{AQ} + \theta_{BQ} - 2\theta_{AQ}\theta_{BQ},$$

$$\theta_{AQ} = \frac{\theta_{BQ} - \theta_{AB}}{2\theta_{BQ} - 1},$$

and

$$\theta_{BQ} = \frac{\theta_{AQ} - \theta_{AB}}{2\theta_{AQ} - 1}.$$

The recombination frequency between the marker loci (θ_{AB}) is estimated by usual methods (Ott 1990) and thereafter fixed, and θ_{AQ} and θ_{BQ} are fixed for some distance from the markers, then the other parameters (the means of the QTL genotypes or linear functions of those means) are estimated. Recombination frequencies between marker and quantitative trait loci are factored out of the model by doing this. Interval mapping is done by estimating test statistics (likelihood ratios or F-statistics) for different θ_{AQ} , thereby fixing θ_{BQ} as well, and finding the θ_{AQ} and θ_{BQ} where the test statistic is maximized. This is grid searching the parameter space of θ_{AQ} and θ_{BQ} . For example, statistics can be estimated for one cM increments of $\hat{\theta}_{AB}$ starting at marker locus A and ending at marker locus B ($0 \leq \theta_{AQ} \leq \hat{\theta}_{AB}$) where $\theta_{BQ} = \frac{\theta_{AQ} - \hat{\theta}_{AB}}{2\theta_{AQ} - 1}$

and $\hat{\theta}_{AB}$ is the maximum likelihood estimate of the recombination frequency between A and B.

The parameters of the genetic model (Table 2) can be estimated without factoring out the recombination frequencies (Knapp et al. 1990; Weller 1993).

Instead of estimating QTL genotype means and test statistics for different recombination frequencies by grid searching (fixing θ_{AQ} and θ_{BQ} for several fixed points between A and B), the recombination frequencies and QTL genotype means can be estimated by minimizing or maximizing some statistic, e.g., by maximizing the likelihood or minimizing the error sum of squares. The genetic model (expected means of marker phenotypes) becomes linear when the recombination frequencies are fixed, and linear least squares methods can then be used to estimate the QTL genotype means. Differences between the various parameter estimation methods are mostly a matter of mechanics. The gain achieved by using linear models, least squares, and grid searching is that virtually any experiment and environment design can be handled using standard software.

The expected means of marker phenotypes (Table 2) can be redefined to implement LIM. Let $\theta_{BQ} = \frac{\theta_{AQ} - \hat{\theta}_{AB}}{2\theta_{AQ} - 1}$, then the expected means of the marker phenotypes are

$$E(\mu_{AB/AB}) = \frac{(1 - \theta_{AB} - 2\theta_{AQ} + \theta_{AQ}\theta_{AB} + \theta_{AQ}^2)\mu_{QQ} + (\theta_{AQ}\theta_{AB} - \theta_{AQ}^2)\mu_{qq}}{(\theta_{AB} - 1)(2\theta_{AQ} - 1)}$$

$$= p_1\mu_{QQ} + p_2\mu_{qq} = \frac{(1 - \theta_{AB} - 2\theta_{AQ} + 2\theta_{AQ}^2)\alpha}{(\theta_{AB} - 1)(2\theta_{AQ} - 1)} = x_1\alpha$$

$$E(\mu_{Ab/aB}) = \frac{(\theta_{AQ} + \theta_{AQ}\theta_{AB} - \theta_{AB} - \theta_{AQ}^2)\mu_{QQ} + (\theta_{AQ}\theta_{AB} - \theta_{AQ} + \theta_{AQ}^2)\mu_{qq}}{2\theta_{AB}\theta_{AQ} - \theta_{AB}}$$

$$= p_3\mu_{QQ} + p_4\mu_{qq} = \frac{-(\theta_{AB} - 2\theta_{AQ} + 2\theta_{AQ}^2)\alpha}{2\theta_{AB}\theta_{AQ} - \theta_{AB}} = -x_2\alpha$$

$$E(\mu_{aB/aB}) = \frac{(\theta_{AQ}\theta_{AB} - \theta_{AQ} + \theta_{AQ}^2)\mu_{QQ} + (\theta_{AQ} + \theta_{AQ}\theta_{AB} - \theta_{AB} - \theta_{AQ}^2)\mu_{qq}}{2\theta_{AB}\theta_{AQ} - \theta_{AB}}$$

$$= p_4\mu_{QQ} + p_3\mu_{qq} = \frac{(\theta_{AB} - 2\theta_{AQ} + 2\theta_{AQ}^2)\alpha}{2\theta_{AB}\theta_{AQ} - \theta_{AB}} = x_2\alpha$$

and

$$E(\mu_{ab/ab}) = \frac{(\theta_{AQ}\theta_{AB} - \theta_{AQ}^2)\mu_{QQ} + (1 - \theta_{AB} - 2\theta_{AQ} + \theta_{AQ}\theta_{AB} + \theta_{AQ}^2)\mu_{qq}}{(\theta_{AB} - 1)(2\theta_{AQ} - 1)}$$

$$= p_2\mu_{QQ} + p_1\mu_{qq} = \frac{-(1 - \theta_{AB} - 2\theta_{AQ} + 2\theta_{AQ}^2)\alpha}{(\theta_{AB} - 1)(2\theta_{AQ} - 1)} = -x_1\alpha \quad (4)$$

where $\mu_{QQ} = \alpha$, $\mu_{qq} = -\alpha$, $\mu_{QQ} - \mu_{qq} = 2\alpha$ and α is the additive effect of the QTL. Substituting $\hat{\theta}_{AB}$ for θ_{AB} and $\hat{\theta}_{AQ}$ for θ_{AQ} , the expected frequencies of the QTL can be estimated where $\hat{\theta}_{AQ}$ is the estimated (fixed) distance of Q from A ($0 \leq \hat{\theta}_{AQ} \leq \hat{\theta}_{AB}$).

The expected means of marker genotypes for $\hat{\theta}_{AB} = 0.10$, $\hat{\theta}_{AQ} = 0.0527864$, and $\hat{\theta}_{BQ} = \frac{0.0527864 - 0.10}{2(0.0527864) - 1} = 0.0527864$, for example, are

$$\begin{aligned} E(\mu_{AB/AB}) &= p_1\mu_{QQ} + p_2\mu_{qq} = 0.9969\mu_{QQ} + 0.0031\mu_{qq} = 0.9938\alpha, \\ E(\mu_{Ab/Ab}) &= p_3\mu_{QQ} + p_4\mu_{qq} = 0.5\mu_{QQ} + 0.5\mu_{qq} = 0.0\alpha, \\ E(\mu_{aB/aB}) &= p_4\mu_{QQ} + p_3\mu_{qq} = 0.5\mu_{QQ} + 0.5\mu_{qq} = 0.0\alpha, \end{aligned}$$

and

$$E(\mu_{ab/ab}) = p_2\mu_{QQ} + p_1\mu_{qq} = 0.0031\mu_{QQ} + 0.9969\mu_{qq} = -0.9938\alpha; \quad (5)$$

thus, the expected means of marker genotypes, with the recombination frequencies fixed, are linear functions of the means of QTL genotypes.

A QTL is mapped by estimating the means of QTL genotypes for $0 \leq \hat{\theta}_{AQ} \leq \hat{\theta}_{AB}$, and testing the hypothesis of no QTL ($H_0: \mu_{QQ} = \mu_{qq}$). Evidence for a QTL between A and B exists when the null hypothesis is rejected for some θ_{AQ} .

QTL parameters can be estimated and hypotheses about differences between QTL genotype means can be tested by using linear contrasts between marker means where the contrast coefficients are defined for fixed θ_{AQ} and θ_{BQ} and marker phenotypes *per se* are used as independent variables. A linear model for the RCB experiment is

$$y_{ikj} = \mu + b_i + m_k + g(m)_{kj} + e_{ikj} \quad (6)$$

where y_{ikj} is the ikj th observation of the quantitative trait, μ is the population mean, b_i is the effect of the i th block, m_k is the effect of the k th marker genotype, and e_{ikj} is the random error for the j th line of the k th marker genotype in the i th block (Table 3). The effects of marker genotypes are fixed, while the effects of other factors are random. The parameters of (6) can be estimated by using standard linear model methods since the marker phenotypes are known. The QTL effects of (2) can be estimated by using linear differences among the marker means of (6). The additive effect of the QTL for doubled haploid lines, for example, is estimated by

Table 3. Degrees of freedom, Type III sum of squares, and expected mean squares for lines tested in randomized complete blocks in one environment where the effects of marker genotypes are fixed and the effects of lines and blocks are random and factors other than marker genotypes are balanced.

Factor	Degrees of freedom ^a	Sum of squares	Expected mean square ^b
Block	$df_R = b - 1$	$R[b \mu, g]$	$E(M_B) = \sigma_E^2 + N\sigma_B^2$
Line (G)	$df_G = N - 1$	$R[g \mu, b]$	$E(M_G) = \sigma_E^2 + r\sigma_G^2$
Marker (M)	$df_M = m - 1$	$R[m \mu, b]$	$E(M_M) = \sigma_E^2 + r\sigma_{G_M}^2 + \phi_M^2$
G:M	$df_{G_M} = N - m$	$R[g(m) \mu, m, b]$	$E(M_{G_M}) = \sigma_E^2 + r\sigma_{G_M}^2$
Residual	$df_E = (N-1)(b-1)$	$R[e \mu, b, g]$	$E(M_E) = \sigma_E^2$

^a m is the number of marker genotypes, $N = \sum_{i=1}^q n_i$ is the number of lines where n_i is the number of lines of the i th marker genotype, and b is the number of blocks.

^b σ_E^2 is the error variance, $\sigma_{G_M}^2$ is the line nested in QTL genotype variance, σ_G^2 is the between line variance, and θ_M^2 is the variance of fixed effects of marker genotypes.

$$2\hat{\alpha} = \hat{x}_A \hat{\mu}' = [\hat{x}_1 \quad -\hat{x}_2 \quad \hat{x}_2 \quad -\hat{x}_1] \begin{bmatrix} \hat{\mu}_{AB/AB} \\ \hat{\mu}_{Ab/Ab} \\ \hat{\mu}_{aB/aB} \\ \hat{\mu}_{ab/ab} \end{bmatrix} = \hat{x}_1 \hat{\mu}_{AB/AB} - \hat{x}_2 \hat{\mu}_{Ab/Ab} + \hat{x}_2 \hat{\mu}_{aB/aB} - \hat{x}_1 \hat{\mu}_{ab/ab}$$

where $\mu = [\mu_{AB/AB} \mu_{Ab/Ab} \mu_{aB/aB} \mu_{ab/ab}]$ is the vector of marker phenotype means and \hat{x}_A is the vector of coefficients for estimating the additive effect of the QTL – the coefficients are estimated by fixing the recombination frequencies.

The hypothesis of no difference between QTL genotype means can be tested by using the F-statistic $F = M_Q/M_{G_M}$ since

$$\frac{E(M_Q)}{E(M_{G_M})} = \frac{\sigma_E^2 + r\sigma_{G_M}^2 + \phi_Q^2}{\sigma_E^2 + r\sigma_{G_M}^2}$$

where M_Q is the mean square for the hypothesis being tested, which is estimated by a difference among marker phenotype means (Table 3), and the null hypothesis is rejected with a Type I error probability of α when $F > F_{\alpha; df_Q, df_{G_M}}$. For doubled haploid lines and one QTL, the null hypothesis is $H_0: \mu_{QQ} = \mu_{qq}$. Because the other parameters (θ_{AB} , θ_{AQ} , and θ_{BQ}) are fixed, the sum of squares for the single degree of freedom additive effect contrast for doubled haploid lines is equal to the sum of squares for marker genotypes, which has three degrees of freedom, as long as the genetic model for the hypothesized QTL is adequate. This works out this way for any mating design where two genotypes are observed at a given QTL, e.g., backcross, recombinant inbred, and various testcross progeny. When one QTL is hypothesized to lie between two segregating marker loci A and B, there are three genotypes at each QTL, and two degrees of freedom for differences between QTL genotype means for F_2 , F_3 , F_4 , or other segregating generations among inbred lines. These can be estimated as linear differences between marker phenotype means, for which there is eight degrees of freedom.

A LIM example is developed for one marker-bracket (*Plc* and *iABI151*)

(Kleinhofs et al. 1992) and an experiment where 150 barley doubled haploid lines were tested in randomized complete blocks experiment designs at Corvallis, Oregon and Pullman, Washington in 1991 (Hayes et al. 1993). The quantitative trait used for the example is seed yield (kg/ha). The number of replications of lines was not balanced. Fifty lines were replicated twice, while the other 100 lines were unreplicated, so the coefficients for the expected mean squares are more complicated than those described above for (6) (Table 3).

The hypothesis of no differences between marker genotype means ($H_0: \mu_{AB/AB} = \mu_{Ab/Ab} = \mu_{aB/aB} = \mu_{ab/ab}$) from experiments within each environment were tested using $F = M_M/M_{G:M}$ where

$$\frac{E(M_M)}{E(M_{G:M})} = \frac{\sigma_E^2 + 1.2\sigma_{G:M}^2 + \phi_M^2}{\sigma_E^2 + 1.3\sigma_{G:M}^2}$$

(Tables 4 and 5). The coefficients for $\sigma_{G:M}^2$ of M_M and $M_{G:M}$ are not equal because the number of replications of lines are unequal. Nor are the Type I and Type III sum of squares equal, and Type III sum of squares must be used to get unbiased test statistics and parameter estimates for this problem (Searle 1971). The mean square for marker genotypes, for example, was estimated by

$$\frac{R[m|\mu, b]}{m - 1}$$

where $R[m|\mu, b]$ is the reduction in sum of squares due to fitting the effects of marker genotypes after the mean and the effects blocks (Tables 4 and 5).

The hypothesis of no differences between marker genotypes *per se* is not important, but since the marker phenotypes are known within errors of ascertainment (misscored marker phenotypes), the coefficients for the mean squares and Type III sum of squares can be directly estimated and used for Type III tests of hypotheses of differences between means of hypothesized QTL genotypes. The hypothesis of no difference between QTL genotype means ($H_0: \mu_{QQ} = \mu_{qq}$) from (6) was tested using $F = M_Q/M_{G:M}$ where

$$\frac{E(M_Q)}{E(M_{G:M})} = \frac{\sigma_E^2 + 1.2\sigma_{G:M}^2 + \phi_Q^2}{\sigma_E^2 + 1.3\sigma_{G:M}^2},$$

ϕ_Q^2 is the variance of fixed effects of QTL genotypes, $M_Q = \frac{R[q|\mu, b]}{q - 1}$ is the mean square for the hypothesis, and $R[q|\mu, b]$ is the reduction in sum of squares for QTL genotypes (estimated as differences between marker phenotype means) estimated after the population mean and the effects of blocks (Tables 4 and 5).

Parameters and test statistics were estimated for every cM between *Plc* and *iABI151* (Fig. 1). The recombination frequency between *Plc* and *iABI151* was $\hat{\theta}_{AB} = 0.18$. Evidence for a QTL between *Plc* and *iABI151* for seed yield was found for lines tested at Corvallis, but not at Pullman (Fig. 1 and Tables 4 and

Table 4. Degrees of Freedom (DF), Type III mean squares (MS), F-statistics (F), and probabilities for F-statistics ($\text{Pr} > F$) for different factors affecting the seed yields (kg/ha) of barley doubled haploid lines tested at Corvallis, Oregon.

Factor	DF	MS	F	Pr>F	Expected mean square ^a
Block	1	6,179,351	5.2	0.03	$E(M_B) = \sigma_E^2 + 50.0\sigma_B^2$
Line (G)	149	2,143,979	1.8	0.01	$E(M_G) = \sigma_E^2 + 1.3\sigma_G^2$
Marker	3	5,764,777	2.7	0.048	$E(M_M) = \sigma_E^2 + 1.2\sigma_{G:M}^2 + \phi_M^2$
QTL (Q)	1	16,559,393	7.8	0.0059	$E(M_Q) = \sigma_E^2 + 1.2\sigma_{G:M}^2 + \phi_Q^2$
G:M	146	2,113,975	1.8	0.011	$E(M_{G:M}) = \sigma_E^2 + 1.3\sigma_{G:M}^2$
Error	49	1,191,234			$E(M_E) = \sigma_E^2$

^a σ_E^2 is the error variance, $\sigma_{G:M}^2$ is the line nested in marker genotype variance, σ_G^2 is the between line variance, ϕ_M^2 is the variance of fixed effects of marker genotypes, and ϕ_Q^2 is the variance of fixed effects of QTL genotypes estimated using differences between marker genotype means.

Table 5. Degrees of freedom (DF), Type III mean squares (MS), F-statistics (F), and probabilities for F-statistics ($\text{Pr} > F$) for different factors affecting the seed yields (kg/ha) of barley doubled haploid lines tested at Pullman, Washington.

Factor	DF	MS	F	Pr>F	Expected mean square ^a
Block	1	1,599,643	2.3	0.14	$E(M_B) = \sigma_E^2 + 50.0\sigma_B^2$
Line (G)	149	1,662,027	2.4	0.0004	$E(M_G) = \sigma_E^2 + 1.3\sigma_G^2$
Marker (M)	3	614,284	0.4	0.75	$E(M_M) = \sigma_E^2 + 1.2\sigma_{G:M}^2 + \phi_M^2$
QTL	1	774,525	0.5	0.48	$E(M_Q) = \sigma_E^2 + 1.2\sigma_{G:M}^2 + \phi_Q^2$
G:M	146	1,689,607	2.4	0.0003	$E(M_{G:M}) = \sigma_E^2 + 1.3\sigma_{G:M}^2$
Error	49	703,162			$E(M_E) = \sigma_E^2$

^a σ_E^2 is the error variance, $\sigma_{G:M}^2$ is the line nested in marker genotype variance, σ_G^2 is the between line variance, ϕ_M^2 is the variance of fixed effects of marker genotypes, and ϕ_Q^2 is the variance of fixed effects of QTL genotypes estimated using differences between marker genotype means.

5). F-statistics for $H_0: \mu_{QQ} \neq \mu_{qq}$ for some $\hat{\theta}_{AQ}$ were significantly greater than $F_{0.01; df_Q, df_{G:M}}$ for lines tested at Corvallis, but not at Pullman. The maximum difference between QTL genotype means was found for $\hat{\theta}_{AQ} = 0.09$ for Corvallis and for $\hat{\theta}_{AQ} = 0.07$ for Pullman, although the latter was not significant (Fig. 1 and Tables 4 and 5). The contrast coefficients and least square means (kg/ha) for $\hat{\theta}_{AQ} = 0.09$ for Corvallis are

$$\hat{x}_A = [0.976 \quad -0.110 \quad 0.110 \quad -0.976]$$

and

$$\hat{\mu} = [\hat{\mu}_{AB/AB} \quad \hat{\mu}_{Ab/Ab} \quad \hat{\mu}_{aB/aB} \quad \hat{\mu}_{ab/ab}] = [6,223.3 \quad 6,407.9 \quad 5,913.0 \quad 5,589.0];$$

so the additive effect of the hypothesized QTL affecting the seed yield of lines tested at Corvallis is $(\hat{x}_A \hat{\mu}')/2 = 168.4$ where $\mu_{QQ} - \mu_{qq} = 2\alpha$.

The parameters and test statistics for this example were estimated using PROC GLM of SAS (1992). The code for the example is

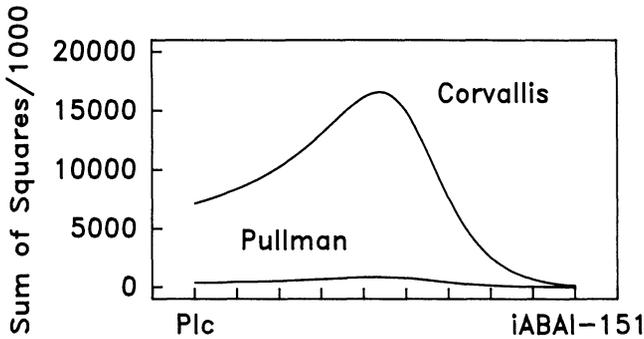


Fig. 1. Seed yield (kg/ha) sum of squares for the additive effect of a QTL between the *Plc* and *iABAI151* locus for barley doubled haploid lines tested in Corvallis, Oregon and Pullman, Washington.

```
proc glm;
  class block line;
  model y = block line;
proc glm;
  class block line marker;
  model y = block marker line (marker);
  contrast 'QTL A' marker 0.976 0.110 -0.110 -0.976;
  lsmeans marker; ;
```

The parameters of (1) were estimated by the first `proc glm` sequence, while the parameters of (6) were estimated by the second `proc glm` sequence. The first `proc glm` sequence is not essential because the between line sum of squares can be estimated by summing the sum of squares for markers and lines nested in markers. The `contrast` statement estimates the sum of squares for the additive effect of the QTL, while the `ls-means` statement estimates the least square means for marker phenotypes.

The versatility of LIM should be clear from this example. Only the `model` statement needs to be rewritten for different experiment designs. The other statements (`marker`, `contrast`, `lsmeans`) are determined by the mating design, and are not affected by different experiment designs.

The problem of estimating QTL parameters from different experiment designs or any experiment design can be simplified by analyzing the least square means of lines instead of the original observations – if the experiment design is balanced, then the least square and arithmetic means of lines are equal (Searle 1971). The experiment is first analyzed ignoring marker genotypes, but not other factors, and the least square means of lines are output and used for subsequent QTL analyses. The standard analysis of lines is done for the experiment design used. By using the least square means of lines for the QTL analysis, observations from different experiment designs can be merged, and one QTL analysis can be used for any experiment design. Pure error is eliminated from the model and the error for testing hypotheses about QTL

Table 6. Degrees of freedom, Type III sum of squares, and expected mean squares for the analysis of the least square means of lines tested in one environment where the effects of marker genotypes are fixed and the effects of lines are random and factors other than marker genotypes are balanced.

Factor	Degrees of freedom ^a	Sum of squares	Expected mean square ^b
Marker (M)	$df_M = m - 1$	$R[m \mu, b]$	$E(M_M) = \sigma_{GM}^2 + \phi_M^2$
Line:M (G:M)	$df_{GM} = N - m$	$R[g(m) \mu, g, b]$	$E(M_{GM}) = \sigma_{GM}^2$

^a m is the number of marker genotypes and $N = \sum_{i=1}^q n_i$ is the number of lines where n_i is the number of lines of the i th marker genotype.

^b σ_{GM}^2 is the line nested in QTL genotype variance and ϕ_M^2 is the variance of fixed effects of marker genotypes.

genotypes is the line nested in QTL genotype variance (Table 7). A line-means analysis can of course be used with methods other than LIM, e.g., with maximum likelihood methods.

Software which handles only one-factor linear models, e.g., MAPMAKER-QTL (Lander and Botstein 1990), can be used to estimate QTL parameters from different experiment designs by inputting the least square means of lines, rather than by inputting the original observations, and ignoring the experiment design. These two choices are different. The former eliminates pure error variance, while the latter confounds the pure error and line nested in marker or QTL

Table 7. Degrees of freedom, Type III sum of squares, and expected mean squares for lines tested in randomized complete blocks in different locations where the effects of QTL genotypes are fixed, the effects of lines, blocks, and locations are random, and factors other than QTL genotypes are balanced.

Factor	Degrees of freedom ^a	Sum of squares	Expected mean squares ^b
Block	$df_B = b - 1$		$E(M_B) = \sigma_E^2 + lN\sigma_B^2$
Locations (L)	$df_L = l - 1$		$E(M_L) = \sigma_E^2 + r\sigma_{GL}^2 + rN\sigma_L^2$
Line (G)	$df_G = N - 1$	$R[g \mu, l, b, g]$	$E(M_G) = \sigma_E^2 + r\sigma_{GL}^2 + r\sigma_G^2$
QTL	$df_Q = q - 1$	$R[q \mu, l, b, g]$	$E(M_Q) = \sigma_E^2 + r\sigma_{GL}^2 + r\sigma_G^2 + r\bar{n}\sigma_{QL}^2 + \phi_Q^2$
G:QTL	$df_{GQ} = N - q$	$R[g(q) \mu, l, b, g, q]$	$E(M_{GQ}) = \sigma_E^2 + r\sigma_{GL}^2 + r\sigma_G^2 + r\sigma_{QL}^2$
G × L	$df_{GL} = (N-1)(l-1)$	$R[gl \mu, l, b, g]$	$E(M_{GL}) = \sigma_E^2 + r\sigma_{GL}^2$
QTL × L	$df_{QL} = (q-1)(l-1)$	$R[ql \mu, l, b, g]$	$E(M_{QL}) = \sigma_E^2 + r\sigma_{GL}^2 + r\bar{n}\sigma_{QL}^2$
G:QTL × L	$df_{GQL} = (N-q)(l-1)$	$R[g(ql) \mu, l, b, g, ql]$	$E(M_{GQL}) = \sigma_E^2 + r\sigma_{GL}^2$
Residual	$df_E = (Nl-1)(b-1)$	$R[e \mu, l, b, g, gl]$	$E(M_E) = \sigma_E^2$

^a q is the number of QTL genotypes, $N = \sum_{i=1}^q n_i$ is the number of lines where n_i is the number of lines of the i th QTL genotype, b is the number of blocks, and l is the number of locations.

^b σ_E^2 is the error variance, σ_G^2 is the between line variance, ϕ_Q^2 is the variance of fixed effects of QTL genotypes, σ_{GQ}^2 is the line nested in QTL genotype variance, σ_{GL}^2 is the line by location variance, σ_{QL}^2 is the QTL genotype by location variance, σ_{GQL}^2 is the line nested in QTL genotype by location variance, and

$$\bar{n} = \frac{N - \sum n_i^2}{q - 1}$$

genotypes variances, and can be less powerful. Whether or not it is less powerful is a function of the number of QTL affecting the trait and how their effects are estimated (Knapp et al. 1993).

Most experiment, mating, and environment designs can be handled using PROC GLM or other software for unbalanced linear models since the only unique parts of estimation are the coefficients for testing hypotheses about differences between means of 'hypothesized QTL genotypes', and these coefficients are not affected by the experiment or environment design. Another way to implement LIM is to use these coefficients as covariates. SAS code for implementing LIM this way is

```
proc glm;
  class block marker line;
  model y = block x marker line(marker);
```

where x is the vector of coefficients for a stated θ_{AB} and θ_{AQ} . Because x is listed in the model statement, but not in the class statement, SAS uses it as a covariate. Hypotheses about QTL genotype differences are tested using Type I mean squares – the covariate absorbs 100% of the between marker genotype sums of squares when θ_{AQ} maximizes the sums of squares for the QTL effects and the data are balanced.

The estimation of covariate coefficients can be built in without difficulty. SAS code for the doubled haploid example, for example, is

```
data a;
input locusa locusb;
cards;
1 1
1 2
2 1
2 2
;
data b;
set a;
rab = 0.18;
raq = 0.09;
rbq = (raq - rab) / (2*raq - 1);
p1 = 0;
p2 = 0;
marker = 0;
if locusa = 1 and locusb = 1 then do;
  p1 = (1 - raq - rbq + raq*rbq) / (1 - raq - rbq +
    2*raq*rbq);
  p2 = (raq*rbq) / (1 - raq - rbq + 2*raq*rbq);
  marker = 1;
end;
else if locusa = 1 and locusb = 2 then do;
  p1 = (rbq - raq*rbq) / (raq + rbq - 2*raq*rbq);
  p2 = (raq - raq*rbq) / (raq + rbq - 2*raq*rbq);
  marker = 2;
```

```

end;
else if locusa = 2 and locusb = 1 then do;
  p1 = (raq - raq*rbq) / (raq + rbq - 2*raq*rbq);
  p2 = (rbq - raq*rbq) / (raq + rbq - 2*raq*rbq);
  marker = 3;
end;
else if locusa = 2 and locusb = 2 then do;
  p1 = (raq*rbq) / (1 - raq - rbq + 2*raq*rbq);
  p2 = (1 - raq - rbq + raq*rbq) / (1 - raq - rbq +
    2*raq*rbq);
  marker = 4;
end;
x = p1 - p2;
output;
data b;
set b;
keep locusa locusb marker x;
proc sort data = b;
by locusa locusb;
data c;
infile 'barley.dat';
input line locusa locusb block y;
proc sort data = c;
by locusa locusb;
data d;
merge b c;
by locusa locusb;
proc print data = d;
proc glm data = d;
  class block marker line;
  model y = block x marker line(marker);
run;

```

The estimation of the covariate (x) uses the expected frequencies defined by (4).

LIM creates a straightforward means for handling replicated progeny and the diversity of linear models arising in plant breeding experiments. This is further exemplified by multiple environment experiments, which are reviewed next.

3. Multiple environment experiments

LIM for multiple environment experiments is handled no differently than for an experiment in one environment. Once coefficients are defined for testing hypotheses about QTL genotypes for a mating design, hypotheses about those genotypes can be tested from a virtually endless number of experiment and environment designs. An example is developed for an experiment done across locations, but it could just as easily be done for an experiment done across years, or for any other environment design. Staying with the RCB experiment design, a linear model for lines tested across locations is

$$y_{ijm} = \mu + b_i + g_j + l_m + gl_{jm} + e_{ijm} \quad (7)$$

where y_{ijm} is the ijm th observation of the quantitative trait, μ is the population mean, b_i is the effect of the i th block, g_j is the effect of the j th line, l_m is the effect of the m th location, $g_j l_m$ is the effect of the interaction between the j th line and the m th location, and e_{ijm} is the random error for the j th line in the m th location in the i th block, while a linear model for QTL genotypes tested across locations is

$$y_{ikmj} = \mu + b_i + q_k + g(q)_{kj} + l_m + ql_{km} + g(ql)_{kmj} + e_{ikmj} \quad (8)$$

where y_{ikmj} is the $ikmj$ th observation of the quantitative trait, μ is the population mean, b_i is the effect on the i th block, q_k is the effect of the k th QTL genotype, $g(q)_{kj}$ is the effect of the j th line nested in the k th QTL genotype, l_m is the effect of the m th location, ql_{km} is the effect of the k th QTL genotype in the m th location, $g(ql)_{kmj}$ is the effect of the j th line nested in the k th QTL genotype in the m th location, and e_{ikmj} is the random error for the j th line nested in the k th QTL genotype in the m th location and the i th block (Table 7).

The sum of squares for lines is the sum of the sum of squares for QTL genotypes and lines nested in QTL genotypes, as before, and the sum of the sum of squares for lines by locations is the sum of the sum of squares for QTL genotypes by locations and for lines nested in QTL genotypes by locations; thus,

$$\begin{aligned} E(M_G) &= \sigma_E^2 + r\sigma_{GL}^2 + rl\sigma_G^2 = \frac{df_Q E(M_Q) + df_{G_Q} E(M_{G_Q})}{df_Q + df_{G_Q}} \\ &= \sigma_E^2 + r\sigma_{G_{QL}}^2 + \frac{r\bar{n}(q-1)}{N-1} \sigma_{QL}^2 + rl\sigma_{G_Q}^2 + \frac{rl\bar{n}}{N-1} \left[\frac{\Sigma q_k^2}{q-1} \right] \end{aligned} \quad (9)$$

and

$$\begin{aligned} E(M_{GL}) &= \sigma_E^2 + r\sigma_{GL}^2 = \frac{df_{QL} E(M_{QL}) + df_{G_{QL}} E(M_{G_{QL}})}{df_{QL} + df_{G_{QL}}} \\ &= \sigma_E^2 + r\sigma_{G_{QL}}^2 + \frac{r\bar{n}(q-1)}{N-1} \sigma_{QL}^2 \end{aligned} \quad (10)$$

where $\sigma_G^2 = \sigma_{G_Q}^2 + \frac{\bar{n}}{N-1} \left[\frac{\Sigma q_k^2}{q-1} \right] = \sigma_{G_Q}^2 + \phi_Q^2$ and $\sigma_{GL}^2 = \sigma_{G_{QL}}^2 + \frac{\bar{n}(q-1)}{N-1} \sigma_{QL}^2$.

Since QTL genotypes are not 'observed', we proceed as before by defining a linear model where differences between marker genotype means are used to estimate QTL parameters. This is done by substituting the effects of QTL genotypes in (8) with the effects of marker genotypes, viz.

$$y_{ikmj} = \mu + b_i + m_k + g(m)_{kj} + l_m + ml_{km} + g(ml)_{kmj} + e_{ikmj}$$

where y_{ikmj} is the $ikmj$ th observation of the quantitative trait, μ is the population mean, b_i is the effect of the j th line nested in the k th marker

Table 8. Degrees of freedom, Type III sum of squares, and expected mean squares for lines tested in randomized complete blocks across locations where the effects of QTL genotypes are fixed, the effects of lines, blocks, and locations are random, and factors other than QTL genotypes are balanced.

Factor	Degrees of freedom ^a	Sum of squares	Expected mean squares ^b
Block	$df_B = b-1$		$E(M_B) = \sigma_E^2 + lN\sigma_B^2$
Locations (L)	$df_L = l-1$		$E(M_L) = \sigma_E^2 + r\sigma_{GL}^2 + rN\sigma_L^2$
Line (G)	$df_G = N-1$	$R[g \mu, l, b, gl]$	$E(M_G) = \sigma_E^2 + r\sigma_{GL}^2 + r\sigma_G^2$
Marker (M)	$df_M = m-1$	$R[m \mu, l, b, gl]$	$E(M_M) = \sigma_E^2 + r\sigma_{GM}^2 + r\sigma_G^2 + r\bar{n}\sigma_{ML}^2 + \phi_M^2$
G:M	$df_{GM} = N-m$	$R[g(m) \mu, l, b, gl, m]$	$E(M_{GM}) = \sigma_E^2 + r\sigma_{GM}^2 + r\sigma_G^2$
G × L	$df_{GL} = (N-1)(l-1)$	$R[gl \mu, l, b, g]$	$E(M_{GL}) = \sigma_E^2 + r\sigma_{GL}^2$
M × L	$df_{ML} = (m-1)(l-1)$	$R[ml \mu, l, b, g]$	$E(M_{ML}) = \sigma_E^2 + r\sigma_{ML}^2 + r\bar{n}\sigma_{ML}^2$
G:M × L	$df_{GML} = (N-m)(l-1)$	$R[g(ml) \mu, l, b, g, ml]$	$E(M_{GML}) = \sigma_E^2 + r\sigma_{GML}^2$
Residual	$df_E = (Nl-1)(b-1)$	$R[e \mu, l, b, g, gl]$	$E(M_E) = \sigma_E^2$

^a m is the number of marker genotypes, $N = \sum_{i=1}^q n_i$ is the number of lines of the i th marker genotype, b is the number of blocks, and l is the number of locations.

^b σ_E^2 is the error variance, σ_G^2 is the between line variance, ϕ_M^2 is the variance of fixed effects of marker genotypes, σ_{GM}^2 is the line nested in marker genotype variance, σ_{GL}^2 is the line by location variance, σ_{ML}^2 is the marker genotype by location variance, σ_{GML}^2 is the line nested in marker genotype by location variance, and

$$\bar{n} = \frac{N - \frac{\sum n_i^2}{N}}{q - 1}$$

phenotype, l_m is the effect of the m th location, ml_{km} is the effect of the k th marker genotype in the m th location, $g(ml)_{kjm}$ is the effect of the j th line nested in the k th marker genotype in the m th location, and e_{kjm} is the random error for the j th line nested in the k th marker genotype in the m th location and the i th block (Table 8).

The most important null hypotheses for [11] are $H_0: \sigma_G^2 = 0$ and $H_0: \sigma_{GL}^2 = 0$ and subsets of $H_0: \sigma_{ML}^2 = 0$ and $H_0: \phi_M^2 = 0$, which test hypotheses about QTL genotypes (Table 8). The first two hypotheses are used to determine whether or not there are significant differences between lines and between lines by locations. The second two hypotheses are used to determine whether or not there are significant differences between QTL genotypes and between QTL genotypes by locations. The QTL genotype by location and marker genotype by location effects are random since the effects of locations are random. For fixed environment factors, e.g., fertilization or irrigation rate, the effects of interactions between QTL or marker genotypes and those effects are fixed, with consequent effects on the expected mean squares.

The hypothesis of no differences between marker genotype means ($H_0: \mu_{AB/AB} = \mu_{aB/ab} = \mu_{aB/ab} = \mu_{ab/ab}$) across locations for (11) are tested using

$$F = M_M / (M_{GM} + M_{ML} - M_{GML}) = (M_M + M_{GML}) / (M_{GM} + M_{ML})$$

since

$$\frac{E(M_M)}{E(M_{GM}) + E(M_{ML}) - E(M_{GML})} = \frac{\sigma_E^2 + r\sigma_{GML}^2 + rl\sigma_{GM}^2 + r\bar{n}\sigma_{ML}^2 + \phi_M^2}{\sigma_E^2 + r\sigma_{GML}^2 + rl\sigma_{GM}^2 + r\bar{n}\sigma_{ML}^2}$$

and specific hypotheses about differences between hypothesized QTL genotypes are tested using

$$F = M_Q / (M_{GM} + M_{ML} - M_{GML}) = (M_Q + M_{GML}) / (M_{GM} + M_{ML}) \quad (12)$$

since

$$\frac{E(M_Q)}{E(M_{GM}) + E(M_{ML}) - E(M_{GML})} = \frac{\sigma_E^2 + r\sigma_{GML}^2 + rl\sigma_{GM}^2 + r\bar{n}\sigma_{ML}^2 + \phi_Q^2}{\sigma_E^2 + r\sigma_{GML}^2 + rl\sigma_{GM}^2 + r\bar{n}\sigma_{ML}^2}$$

where M_Q is the mean square for differences between marker genotype means and $E(M_Q)$ is the expected value of M_Q (Table 8). M_Q is estimated using the appropriate contrast or contrasts for the mating design used. These are the sort of complex F-tests which arise in most multiple environment breeding experiments (Gardner 1963). Since (12) is a linear function of sums of mean squares, the error degrees of freedom is not known and must be estimated (Searle 1970). Numerator and denominator degrees of freedom for (12) are estimated by

$$\frac{[M_Q + M_{GML}]^2}{M_Q^2/(df_Q) + M_{GML}^2/(df_{GML})}$$

and

$$\frac{[M_{GM} + M_{ML}]^2}{M_{GM}^2/(df_{GM}) + M_{ML}^2/(df_{ML})},$$

respectively (Searle 1970).

The hypothesis of no differences between marker genotype means by locations or of no differences between QTL genotypes by locations is more straightforward. The latter is tested using

$$F = M_{QL} / M_{GML}$$

since

$$\frac{E(M_{QL})}{E(M_{GML})} = \frac{\sigma_E^2 + r\sigma_{GML}^2 + r\bar{n}\sigma_{QL}^2}{\sigma_E^2 + r\sigma_{GML}^2}$$

(Table 8).

Type III sum of squares must be used to get unbiased hypothesis tests for marker genotypes or QTL genotypes and for marker genotypes by locations or QTL genotypes by locations since marker genotypes are unbalanced and other factors might not be balanced either (Searle 1970). The mean square for marker genotypes is estimated by $\frac{R[m|\mu,l,b,g]}{m-1}$ where $R[m|\mu,l,b,g]$ is the reduction in

sum of squares for marker genotypes estimated after the mean and the effects blocks, locations, and marker genotypes by locations (Table 8).

The LIM steps outlined for one environment are repeated to test hypotheses about QTL genotype effects across environments and QTL genotype by environment effects. It is only necessary to define contrast coefficients for the QTL genotype by environment tests since the coefficients for QTL genotypes across environments are equal to those for individual environments. This is illustrated for the doubled haploid example. The null hypothesis of no additive effect of a QTL across two locations estimated using doubled haploid lines is

$$H_0: \mu_{QQ1} + \mu_{QQ2} = \mu_{qq1} + \mu_{qq2},$$

This effect is estimated by $\hat{\mu}_{QQ1} - \hat{\mu}_{qq1} + \hat{\mu}_{QQ2} - \hat{\mu}_{qq2}$. The null hypothesis of no additive by location effect of a QTL estimated using doubled haploid lines is

$$H_0: \mu_{QQ1} + \mu_{qq2} = \mu_{qq1} + \mu_{QQ2}.$$

This effect is estimated by $\hat{\mu}_{QQ1} - \hat{\mu}_{qq1} - \hat{\mu}_{QQ2} + \hat{\mu}_{qq2}$. To test these hypotheses, we proceed by estimating the parameters of (11), e.g., the sum of squares for each factor and the least square means for marker genotypes and marker genotypes by environments, and by defining linear differences between these means.

The parameters of (11) can be estimated by using standard methods since the marker phenotypes are known. The additive effect of the QTL for doubled haploid lines tested across locations is estimated by

$$2\hat{\alpha} = \hat{x}_A \hat{\mu}' = \begin{bmatrix} \hat{x}_1 & -\hat{x}_2 & \hat{x}_2 & -\hat{x}_1 \end{bmatrix} \begin{bmatrix} \hat{\mu}_{AB/AB} \\ \hat{\mu}_{Ab/Ab} \\ \hat{\mu}_{aB/aB} \\ \hat{\mu}_{ab/ab} \end{bmatrix} = \hat{x}_1 \hat{\mu}_{AB/AB} - \hat{x}_2 \hat{\mu}_{Ab/Ab} + \hat{x}_2 \hat{\mu}_{aB/aB} + \hat{x}_1 \hat{\mu}_{ab/ab}$$

where $\mu = [\mu_{AB/AB} \mu_{Ab/Ab} \mu_{aB/aB} \mu_{ab/ab}]$ are the means of marker phenotypes across locations and \hat{x}_A are the estimated coefficients for estimating the additive effect of the QTL – the coefficients are estimated by fixing the recombination frequencies.

Coefficients for testing the hypothesis of no QTL genotype by environment interaction are found by multiplying coefficients for testing the hypothesis of no difference between environment means by coefficients for testing the hypothesis of no difference between QTL genotype means. Coefficients for testing the hypothesis of no difference between two locations, for example, are $x_L = [1 \ -1]$. Scalar-multiplying this vector by \hat{x}_A we get

Table 9. Degrees of freedom (DF), Type III mean squares (MS), F-statistics (F), probabilities for F-statistics (Pr > F), and expected mean squares (EMS) for different factors affecting the seed yields (kg/ha) of barley doubled haploid lines tested at Corvallis, Oregon and Pullman, Washington.

Factor	DF	MS	F	Pr>F	EMS ^a
Location (L)	1	8,894,799	5.1	0.03	$E(M_L)=\sigma_E^2+1.2\sigma_{GL}^2+180.0\sigma_L^2$
Block	1	7,033,500	7.4	0.008	$E(M_B)=\sigma_E^2+100.0\sigma_B^2$
Line (G)	149	2,054,952	1.2	0.1	$E(M_G)=\sigma_E^2+1.3\sigma_{GL}^2+2.6\sigma_G^2$
Marker (M)	3	2,200,611	0.6	0.8	$E(M_M)=\sigma_E^2+1.2\sigma_{G:ML}^2+2.3\sigma_{G:M}^2+38.5\sigma_{ML}^2+\phi_M^2$
QTL (Q)	1	4,922,170	1.1	0.5	$E(M_Q)=\sigma_E^2+1.2\sigma_{G:ML}^2+2.3\sigma_{G:M}^2+38.5\sigma_{ML}^2+\phi_Q^2$
G:M	146	2,064,989	1.2	0.1	$E(M_{G:M})=\sigma_E^2+1.3\sigma_{G:ML}^2+2.6\sigma_{G:M}^2$
G × L	149	1,741,562	1.8	0.0006	$E(M_{GL})=\sigma_E^2+1.3\sigma_{GL}^2$
M × L	3	2,064,989	1.2	0.3	$E(M_{ML})=\sigma_E^2+1.2\sigma_{G:ML}^2+38.5\sigma_{ML}^2$
Q × L	1	12,514,198	7.3	0.008	$E(M_{QL})=\sigma_E^2+1.2\sigma_{G:ML}^2+38.5\sigma_{QL}^2$
G:M × L	146	1,710,704	1.8	0.0009	$E(M_{G:ML})=\sigma_E^2+1.3\sigma_{G:ML}^2$
Error	99	945,160			$E(M_{G:ML})=\sigma_E^2$

^a σ_E^2 is the error variance, σ_G^2 is the between line variance, ϕ_M^2 is the variance of fixed effects of QTL genotypes, $\sigma_{G:M}^2$ is the line nested in marker genotype variance, σ_{GL}^2 is the line by location variance, σ_{ML}^2 is the marker genotype by location variance, and $\sigma_{G:ML}^2$ is the line nested in marker genotype by location variance.

$$[\hat{x}_1 \ -\hat{x}_2 \ \hat{x}_2 \ -\hat{x}_1 \ -\hat{x}_1 \ \hat{x}_2 \ -\hat{x}_2 \ \hat{x}_1]$$

where

$$[\mu_{AB/AB1} \ \mu_{Ab/Ab1} \ \mu_{aB/aB1} \ \mu_{ab/ab1} \ \mu_{AB/AB2} \ \mu_{Ab/Ab2} \ \mu_{aB/aB2} \ \mu_{ab/ab2}]$$

is the vector of marker phenotype means for the two locations. Coefficients can be defined for any number of environments in this way as shown below.

A marker-bracket is interval mapped as for one environment, but with test-statistics estimated for the effect of QTL genotypes in each environment, the mean effect of QTL genotypes across environments, and the effects of QTL genotype by environment interactions. The θ_{AQ} at which the test-statistics for each of these tests is maximum can be and often is different. The most probable distance of the QTL from either of the marker loci could be estimated by finding the maximum for the pooled hypothesis test.

Since marker genotypes and blocks were unbalanced in the barley example, the variance coefficients for the expected mean squares are functions of the numbers of observations within subsets of cells (Table 9). The analysis is messy but doable. Parameters and test statistics for (11) were estimated for every cM between *Plc* and *iABI151* (Fig. 1). The contrast coefficients and least square means for seed yields of doubled haploid lines across environments (kg/ha) for $\hat{\theta}_{AB} = 0.18$ and $\hat{\theta}_{AQ} = 0.09$ for the *Plc-iABI151* marker-bracket are

$$\hat{x}_A = [0.976 \ -0.110 \ 0.110 \ -0.976]$$

and

$$\hat{\mu} = [\hat{\mu}_{AB/AB} \hat{\mu}_{Ab/Ab} \hat{\mu}_{aB/aB} \hat{\mu}_{ab/ab}] = [5,825.2 \ 6,009.6 \ 5,820.7 \ 5,580.3];$$

thus, the additive effect of the hypothesized QTL between *Plc* and *iABI151* is $(\hat{x}_A \hat{\mu}') / 2 = 129.9$. The hypothesis of no differences between marker genotype means across locations and no additive effect of the QTL across locations were tested by

$$F = (M_M + M_{G_{ML}}) / (M_{G_M} + M_{ML}) = 0.6$$

and

$$F = (M_Q + M_{G_{ML}}) / (M_{G_M} + M_{ML}) = 1.1,$$

respectively, where the estimated numerator degrees of freedom for the former and latter are

$$\frac{[M_M + M_{G_{ML}}]^2}{M_M^2/df_M + M_{G_{ML}}^2/df_{G_{ML}}} = 9.4 \cong 9.0$$

and

$$\frac{[M_Q + M_{G_{ML}}]^2}{M_Q^2/df_Q + M_{G_{ML}}^2/df_{G_{ML}}} = 1.8 \cong 2.0,$$

respectively, and the estimated denominator degrees of freedom for either test are

$$\frac{[M_{G_M} + M_{ML}]^2}{M_{G_M}^2/df_{G_M} + M_{ML}^2/df_{ML}} = 2.2 \cong 2.0$$

(Searle 1970). Differences between marker genotype means were not significant ($0.6 < F_{0.09, 9, 2}$) across locations (Table 9). The additive effect of the QTL across locations was not significant ($1.1 < F_{0.01, 9, 2}$) for any $\hat{\theta}_{AQ}$ (Fig. 1 and Table 9). Nor was there any evidence ($F = M_G / M_{GL} = 1.2$) for genetic variance between lines tested across locations ($1.2 < F_{0.01, 149, 149}$) (Table 9). There was evidence ($F = M_{GL} / M_E = 1.8$) for a line by location interaction ($1.8 > F_{0.001, 149, 99}$) and for a QTL genotype by location interaction ($7.3 > F_{0.001, 1, 146}$) where $F = M_{QL} / M_{G_{ML}} = 7.3$ (Fig. 1 and Table 9).

This example illustrates how QTL parameters can be estimated for a typical multiple environment breeding experiment. The parameters and test statistics for this example were estimated using SAS (1992). The code for the example is

```
proc glm data = b;
  class location block line marker;
  model y = location block marker line(marker)
         location*marker location*line(marker);
  contrast 'QTL A' marker 0.976 0.110 -0.110 -0.976;
```

```
contrast 'QTL A x L' location*marker
  0.976 0.110 -0.110 -0.976 -0.976 -0.110 0.110 0.976;
```

The only difference between this code and the code shown for the analysis of the individual environment experiments, other than the `model` statement, is the addition of the `contrast` statement for testing the hypothesis of no QTL genotype by location interaction. We showed how to get the coefficients earlier.

Additional contrast coefficients are needed if lines are tested in more than two environments. One way to handle any number of environments is to use orthogonal polynomials to define contrast coefficients for environments and to subsequently scalar-multiply the environment coefficient matrix by X_A where X_A is made by stacking e x_A vectors. The scalar product is an $(e - 1) \times 4e$ matrix of coefficients. For three environments, for example, orthogonal polynomials for estimating contrasts between the three environment means, e.g., $[\mu_1 \mu_2, \mu_3]$, are

$$x_E = \begin{bmatrix} 1 & 0 & -1 \\ 1 & -2 & 1 \end{bmatrix},$$

so the coefficients found by scalar multiplying this matrix by

$$\hat{X}_A = \begin{bmatrix} \hat{x}_1 & -\hat{x}_2 & \hat{x}_2 & -\hat{x}_1 \\ \hat{x}_1 & -\hat{x}_2 & \hat{x}_2 & -\hat{x}_1 \end{bmatrix}$$

are

$$\begin{bmatrix} \hat{x}_1 & -\hat{x}_2 & \hat{x}_2 & -\hat{x}_1 & 0 & 0 & 0 & -\hat{x}_1 & \hat{x}_2 & -\hat{x}_2 & \hat{x}_1 \\ \hat{x}_1 & -\hat{x}_2 & \hat{x}_2 & -\hat{x}_1 & -2\hat{x}_1 & 2\hat{x}_2 & -2\hat{x}_2 & 2\hat{x}_1 & \hat{x}_1 & -\hat{x}_2 & \hat{x}_2 & -\hat{x}_1 \end{bmatrix}.$$

Contrast coefficients can be defined for any environment design. For multiple location and year experiments, for example, coefficients can be defined for testing hypotheses about QTL genotype by location by year interactions by using contrasts among marker genotype by location by year means, in addition to those for testing hypotheses about QTL genotype by year and QTL genotype by location interactions.

Suppose, for the final example, lines are tested in randomized complete blocks across locations and years. To estimate the parameters of lines only, the linear model

$$y_{ijmn} = \mu + b_i + g_j + l_m + y_n + gl_{jm} + gy_{jn} + ly_{mn} + gly_{jmn} + e_{ijmn} \quad (13)$$

can be used where y_{ijmn} is the ijm th observation of the quantitative trait, μ is the population mean, b_i is the effect of the i th block, g_j is the effect of the j th line, l_m is the effect of the m th location, y_n is the effect of the n th year, gl_{jm} is the

Table 10. Degrees of freedom, Type III sum of squares, and expected mean squares for lines tested in randomized complete blocks across locations and years where the effects of QTL genotypes are fixed, the effects of lines, blocks, locations, and years are random, and factors other than QTL genotypes are balanced.

Factor	Degrees of freedom ^a	Sum of squares	Expected mean squares ^b
Block	$df_B = b-1$		
Locations (L)	$df_L = l-1$		
Years (Y)	$df_Y = y-1$		
Y × L	$df_{YL} = (l-1)(y-1)$		
Line (G)	$df_G = N-1$	$R[g \mu, l, y, b, g, l, g, y, g, l, y]$	$E(M_G) = \sigma_E^2 + r\sigma_{GLY}^2 + rl\sigma_{GY}^2 + ry\sigma_{GL}^2 + rly\sigma_G^2$
QTL	$df_Q = q-1$	$R[q \mu, l, y, b, q, l, q, y, q, l, y]$	$E(M_Q) = \sigma_E^2 + r\sigma_{QLY}^2 + r\bar{n}\sigma_{QL}^2 + r\bar{n}\sigma_{QY}^2 + rl\sigma_{GY}^2 + ry\bar{n}\sigma_{QL}^2 + rly\sigma_G^2 + rly\bar{n}\phi_Q^2$
G:QTL	$df_{GQ} = N-q$	$R[g:q \mu, l, y, b, q, l, q, y, q, l, y]$	$E(M_{GQ}) = \sigma_E^2 + r\sigma_{QLY}^2 + rl\sigma_{QY}^2 + ry\sigma_{GL}^2 + rly\sigma_G^2 + rly\bar{n}\phi_Q^2$
G × L	$df_{GL} = (N-1)(l-1)$	$R[gl \mu, l, y, b, g, l, g, y, g, l, y]$	$E(M_{GL}) = \sigma_E^2 + r\sigma_{GLY}^2 + ry\sigma_{GL}^2$
QTL × L	$df_{QL} = (q-1)(l-1)$	$R[ql \mu, l, y, b, q, l, q, y, q, l, y]$	$E(M_{QL}) = \sigma_E^2 + r\sigma_{QLY}^2 + r\bar{n}\sigma_{QL}^2 + rly\sigma_G^2 + rly\bar{n}\phi_Q^2$
G:QTL × L	$df_{GQL} = (N-q)(l-1)$	$R[g:ql \mu, l, y, b, q, l, q, y, q, l, y]$	$E(M_{GQL}) = \sigma_E^2 + r\sigma_{QLY}^2 + ry\sigma_{GL}^2$
G × Y	$df_{GY} = (N-1)(y-1)$	$R[gy \mu, l, y, b, g, l, g, y, g, l, y]$	$E(M_{GY}) = \sigma_E^2 + r\sigma_{GLY}^2 + rl\sigma_{GY}^2$
QTL × Y	$df_{QY} = (q-1)(y-1)$	$R[qy \mu, l, y, b, q, l, q, y, q, l, y]$	$E(M_{QY}) = \sigma_E^2 + r\sigma_{QLY}^2 + r\bar{n}\sigma_{QY}^2 + rl\sigma_{GY}^2 + r\bar{n}\phi_Q^2$
G:QTL × Y	$df_{GQY} = (N-q)(y-1)$	$R[g:qy \mu, l, y, b, q, l, q, y, q, l, y]$	$E(M_{GQY}) = \sigma_E^2 + r\sigma_{QLY}^2 + rl\sigma_{QY}^2$
G × L × Y	$df_{GLY} = (N-1)(l-1)(y-1)$	$R[gly \mu, l, y, b, g, l, g, y, g, l, y]$	$E(M_{GLY}) = \sigma_E^2 + r\sigma_{GLY}^2$
QTL × L × Y	$df_{QLY} = (q-1)(l-1)(y-1)$	$R[qly \mu, l, y, b, q, l, q, y, q, l, y]$	$E(M_{QLY}) = \sigma_E^2 + r\sigma_{QLY}^2 + r\bar{n}\sigma_{QLY}^2$
G:QTL × L × Y	$df_{GQLY} = (N-q)(l-1)(y-1)$	$R[g:qly \mu, l, y, b, q, l, q, y, q, l, y]$	$E(M_{GQLY}) = \sigma_E^2 + r\sigma_{QLY}^2$
Residual	$df_E = (Nl-1)(b-1)$		$E(M_E) = \sigma_E^2$

^a q is the number of QTL genotypes, $N = \sum_{i=1}^q n_i$ is the number of lines where n_i is the number of lines of the i th QTL genotype, b is the number of blocks, l is the number of locations, and y is the number of years.

^b σ_E^2 is the error variance, σ_G^2 is the between line variance, ϕ_Q^2 is the variance of fixed effects of QTL genotypes, σ_{GQ}^2 is the line nested in QTL genotype variance, σ_{GL}^2 is the line by location variance, σ_{QL}^2 is the QTL genotype by location variance, σ_{QL}^2 is the line nested in QTL genotype by location variance, σ_{GY}^2 is the line by year variance, σ_{QY}^2 is the QTL genotype by year variance, σ_{GQY}^2 is the line nested in QTL genotype by year variance, σ_{GLY}^2 is the line by location by year variance, σ_{QLY}^2 is the QTL genotype by location by year variance, σ_{QLY}^2 is the line nested in QTL genotype by location by year variance, and

$$\bar{n} = \frac{N - \frac{\sum n_i^2}{N}}{q - 1}$$

effect of the interaction between the j th line and the m th location, gy_{jn} is the effect of the interaction between the j th line and the n th year, ly_{mn} is the effect of the interaction between the m th location and the n th year, gly_{jmn} is the effect of the interaction between the j th line, the m th location, and the n th year, e_{ijmn} is the random error for the j th line in the m th location, n th year, and i th block, and the effects of lines, locations, years, and blocks are random (Table 10).

Rewriting (13) as a function of QTL parameters leads to

$$y_{ijkmn} = \mu + b_i + q_k + g(q)_{kj} + l_m + ql_{km} + g(ql)_{kmj} + y_n + gy_{kn} + g(qy)_{knj} + ly_{mn} + gly_{kmn} + g(qly)_{kmnj} + e_{ijkmn} \quad (14)$$

where y_{ijkmn} is the $ijkm$ th observation of the quantitative trait, μ is the population mean, b_i is the effect of the i th block, q_k is the effect of the k th QTL genotype, $g(q)_{kj}$ is the effect of the j th line nested in the k th QTL genotype, l_m is the effect of the m th location, ql_{km} is the effect of the interaction between the k th QTL genotype and the m th location, $g(ql)_{kmj}$ is the effect of the j th line nested in the k th QTL genotype and the m th location, y_n is the effect of the n th year, gy_{kn} is the effect of the interaction between the k th QTL genotype and the n th year, $g(qy)_{knj}$ is the effect of the j th line nested in the k th QTL genotype and the n th year, ly_{mn} is the effect of the interaction between the m th location and the n th year, gly_{kmn} is the effect of the interaction between the k th QTL genotype, m th location, and n th year, $g(qly)_{kmnj}$ is the effect of the j th line nested in the k th QTL genotype, m th location, and n th year, and e_{ijkmn} is the random error for the j th line in the k th QTL genotype, m th location, n th year, and i th block (Table 10). There are three QTL genotype by environment interaction variances to estimate in addition to the effects of QTL genotypes across environments. The expected mean squares for lines and between lines by environments can be expressed as a function of the QTL genotype and QTL genotype by environment expected mean squares as shown for the other examples. The pooled expected mean square for lines from (13) and (14), for example, is

$$\begin{aligned} E(M_G) &= \sigma_E^2 + r\sigma_{GLY}^2 + rl\sigma_{GY}^2 + ry\sigma_{GL}^2 + rl\sigma_G^2 = \frac{df_Q E(M_Q) + df_{GQ} E(M_{GQ})}{df_Q + df_{GQ}} \\ &= \sigma_E^2 + r\sigma_{GQLY}^2 + \frac{r\bar{n}(q-1)}{N-1} \sigma_{QLY}^2 + rl\sigma_{GQY}^2 + \frac{rl\bar{n}(q-1)}{N-1} \sigma_{QY}^2 + ry\sigma_{GL}^2 \\ &\quad + \frac{ry\bar{n}(q-1)}{N-1} \sigma_{QL}^2 + rly\sigma_{GL}^2 + \frac{rly\bar{n}}{N-1} \left[\frac{\sum q_k^2}{q-1} \right]; \end{aligned}$$

hence,

$$\sigma_{GLY}^2 = \sigma_{GQLY}^2 + \frac{\bar{n}(q-1)}{N-1} \sigma_{QLY}^2, \quad \sigma_{GY}^2 = \sigma_{GQY}^2 + \frac{\bar{n}(q-1)}{N-1} \sigma_{QY}^2,$$

$$\sigma_{GL}^2 = \sigma_{GQL}^2 + \frac{\bar{n}(q-1)}{N-1} \sigma_{QL}^2, \text{ and } \sigma_G^2 \sigma_{GQ}^2 + \frac{\bar{n}}{N-1} \left[\frac{\sum q_k^2}{q-1} \right] = \sigma_{GQ}^2 + \phi_Q^2 \text{ (Table 10).}$$

The F-statistic for testing the hypothesis of no differences between QTL genotype means is

$$F = \frac{M_Q + M_{GQL} + M_{GQY} + M_{GQLY}}{M_{GQ} + M_{QL} + M_{QY} + M_{QLY}}$$

where

$$\begin{aligned} & \frac{E(M_Q) + E(M_{GQL}) + E(M_{GQY}) + E(M_{GQLY})}{E(M_{GQ}) + E(M_{QL}) + E(M_{QY}) + E(M_{QLY})} \\ &= \frac{\sigma_E^2 + r\sigma_{GQLY}^2 + r\bar{n}\sigma_{QLY}^2 + r\bar{n}\sigma_{QY}^2 + r\sigma_{GQY}^2 + ry\bar{n}\sigma_{QL}^2 + ry\sigma_{GQL}^2 + rly\sigma_{GQ}^2 + \phi_Q^2}{\sigma_E^2 + r\sigma_{GQLY}^2 + r\bar{n}\sigma_{QLY}^2 + r\bar{n}\sigma_{QY}^2 + r\sigma_{GQY}^2 + ry\bar{n}\sigma_{QL}^2 + ry\sigma_{GQL}^2 + rly\sigma_{GQ}^2} \end{aligned}$$

(Table 10) with approximate numerator and denominator degrees of freedom

$$\frac{[M_Q + M_{GQL} + M_{GQY} + M_{GQLY}]^2}{M_Q^2/(df_Q) + M_{GQL}^2/(df_{GQL}) + M_{GQY}^2/(df_{GQY}) + M_{GQLY}^2/(df_{GQLY})}$$

and

$$\frac{[M_{GQ} + M_{QL} + M_{QY} + M_{QLY}]^2}{M_{GQ}^2/(df_{GQ}) + M_{QL}^2/(df_{QL}) + M_{QY}^2/(df_{QY}) + M_{QLY}^2/(df_{QLY})},$$

respectively (Satterthwaite 1943; Searle 1970). Complex F-statistics are needed for testing hypotheses about the QTL genotype by location and QTL genotype by year variances as well. This analysis is messy as is, and worsens when additional factors are unbalanced. The most expedient tack might be to analyze lines ignoring marker and QTL parameters, and then to analyze marker phenotypes (QTL) by using line by environment least square means as the dependent variable. Complex F-statistics arise either way, but the model and analysis are simplified and standardized by using line by environment least square means.

4. Replications of lines and replications of QTL genotypes

The sample size of an experiment (the total number of experimental units) is equal to Nr . Increasing r is usually less expensive than increasing N to achieve a given sample size because the marker phenotypes of each line must be assayed,

and this is usually more expensive than increasing the number of replications of lines. The number of replications of QTL genotypes (\bar{n}) is determined by the total number of lines (N) and by the genetic model. The number of replications of lines (r) is fixed for a given experiment, while the number of replications of QTL genotypes (\bar{n}) and how these replications are laid out within blocks or incomplete blocks is determined by the genetic model. The number of replications of QTL genotypes fluctuates as the number of QTL parameters increases or decreases – \bar{n} decreases as the number of QTL parameters increases. Any of several genetic models might be tested, and the number of genetic models which might be tested for a given experiment often exceeds the number of observations for the experiment. The spatial layout of replications of QTL genotypes within blocks differs for every genetic model and from marker-bracket to marker-bracket.

Power for mapping QTL is affected differently by replications of lines and replications of QTL genotypes (Knapp and Bridges 1991). To estimate QTL genotype means and test hypotheses about QTL genotypes, individuals or lines need not be replicated as long as marker or QTL genotypes are replicated (Knapp 1989; Knapp and Bridges 1990). Unreplicated progeny are nevertheless disadvantageous. By using replicated lines, many useful classical quantitative genetic and QTL parameters can be estimated, and a minimum number of lines need to be assayed for their marker phenotypes (for a given overall sample size). In addition, replicated lines are essential for estimating errors for testing hypotheses about QTL genotype effects and QTL genotype by environment interaction effects as shown earlier (Tables 1, 3, 4, 5, 7, 8, 9 and 10).

Increasing the number of replications of lines only increases power for mapping QTL when most of the between line variance has been explained by differences between QTL genotypes (Knapp and Bridges 1990).

$$\hat{\sigma}_{G,Q}^2 \cong 0 \text{ and } \hat{\sigma}_G^2 \cong \frac{\bar{n}}{N-1} \left[\frac{\sum q_k^2}{q-1} \right] \text{ when this is achieved. When a genome is searched}$$

marker-bracket by marker-bracket, the line nested in QTL genotype variance ($\hat{\sigma}_{G,Q}^2$) is rarely if ever close to zero. As model building progresses, more and more QTL are added to the model and more and more QTL parameters are estimated until, ultimately, $\hat{\sigma}_{G,Q}^2 \equiv 0$; whereupon power is greatly increased for testing hypotheses about QTL genotypes by increasing the number of replications of lines (Knapp and Bridges 1990).

This problem is closely tied to model building and greatly impacts the gain which can be achieved from marker-assisted selection (MAS). Gains from MAS are maximized by maximizing the variance explained by marker genotypes (Lande and Thompson 1990) or by QTL genotypes, which is the goal of a QTL mapping experiment (Lander and Botstein 1989; Knapp and Bridges 1990; Knapp et al. 1992; Van Ooijen 1992). When this is achieved, $\hat{\sigma}_{G,Q}^2 \cong 0$ and $\hat{\sigma}_G^2 \cong \hat{\phi}_G^2$ since $\hat{\sigma}_G^2 = \hat{\sigma}_{G,Q}^2 + \hat{\phi}_G^2$. $\hat{\sigma}_{G,Q}^2$ tends to zero when the parameters of the most

important QTL have been estimated, and the coefficient $\frac{\bar{n}(q-1)}{N-1}$ tends to 1.0 as

q increases. With a balanced number of replications of each QTL genotype, which can only happen for certain progeny types and genetic models, $\bar{n} = 1$ and $\sigma_G^2 = \phi_Q^2$ when $N = q$ (Table 1). $\sigma_G^2 = \sigma_{G-Q}^2$ when no QTL are mapped, and $\sigma_{G-Q}^2 = 0$ when all of the important QTL have been mapped.

Every line within a sample could have a unique QTL genotype. Whether or not they do is a function of the number of lines tested (N) and the number of QTL segregating within the population, which can be different from the number of QTL modeled (q). If a great number of QTL underlie a trait, say 10 independent loci, then $3^{10} = 59,049$ genotypes can arise within an F_2 or F_3 population, while $2^{10} = 1,024$ genotypes can arise within a DH or recombinant inbred line population. For the example of 10 independent QTL, each DH, RI, or F_3 line within most samples almost surely has a unique ten locus QTL genotype.

If the QTL genotype of each line is unique, and the parameters of $q = N$ QTL genotypes are estimated, then

$$N - 1 = q - 1, \bar{n} = \frac{N-1}{q-1} = 1, \hat{\sigma}_{GL}^2 = \frac{\bar{n}q - \bar{n}}{N-1} \hat{\sigma}_{QL}^2 = \hat{\sigma}_{QL}^2,$$

and $\hat{\sigma}_G^2 = \hat{\phi}_Q^2$. A complete QTL model explains a maximum of $100H$ percent of the line-mean phenotypic variance ($\hat{\sigma}_p^2$) or 100 percent of the between line variance where H is the family-mean heritability (Table 2).

σ_{G-QL}^2 is the line by location variance which is not explained by the mapped QTL. σ_{G-Q}^2 and σ_{G-QL}^2 are 'lack-of-fit' variances. Some of the QTL underlying significant QTL genotype by environment interaction effects might be different from those underlying significant QTL effects across environments. This has ramifications for searching genomes for important QTL. If the mean effects of QTL across environments are used to determine which QTL are retained in a model, then QTL underlying a line by environment interaction might be overlooked. This can be avoided by using the effects of QTL by environments in addition to the effects of QTL across environments to determine which QTL should be retained in a model.

5. Multiple loci

One challenge for implementing QTL mapping methods for a wide range of mating designs is defining frequencies of marker and quantitative trait loci, linked and unlinked, for several loci, and then grid searching (interval mapping) the multi-dimensional distances between the markers. Joint frequencies are needed to estimate Type III sum of squares and to get unbiased hypothesis tests and parameter estimates (Knapp et al. 1993); however, this poses a special

problem, and ultimately leads to a paradox. The paradox is that once several marker-brackets are added to the model, missing cells can arise (Knapp et al. 1993). Unbiased QTL effects can then only be estimated when there is no interaction between the loci. Since these interactions cannot be estimated when there are missing cells, the only recourse is to estimate the parameters and Type III sum of squares with the knowledge that the estimates could be biased. Although it is hard to imagine much bias caused by interaction (epistasis), the question of the extent of epistasis still needs to be answered, and it very well could be important for a given population or trait. MAS is probably not going to be affected much by epistasis caused bias, but it might be enhanced by jointly estimating the intralocus (main) effects of a maximum number of QTL.

Knapp (1991), Haley and Knott (1992), and Carbonell et al. (1992) gave two-locus examples of joint frequencies for different mating types. The estimation of these frequencies for more than two loci is straightforward but can be very cumbersome. Hypotheses about additive by additive, additive by dominant, and dominant by dominant QTL genotype interactions can be tested by using contrasts between marker genotype means. The coefficients for these contrasts are defined by the joint expected frequencies of the QTL for fixed recombination frequencies.

The computing becomes burdensome as the number of marker-brackets or QTL increases. One problem posed by multilocus estimation is the number of tests needed to interval map more than one QTL simultaneously. The number for k QTL is $n_i + 1 \cdot n_{i+1} + 1 \cdot \dots \cdot n_k + 1$ where n_i is the distance between the i th and $i+1$ marker loci, and statistics are estimated for every 1 cM between two markers.

The problem of estimating the parameters of multiple QTL warrants much further study. Rodolphe and Lefort (1993) and Jansen (1993) examined the problem and proposed methods for estimating the parameters of multiple QTL. The sort of advances they propose are essential for gaining a less biased and more thorough understanding of the genetics of complex traits.

6. Selecting sources of favorable alleles and marker-assisted selection

Two of the goals of QTL mapping experiments are to find sources of favorable alleles for developing superior cultivars and hybrids and to gain the knowledge necessary for maximizing selection gains through marker-assisted selection (MAS) (Tanksley et al. 1989; Lande and Thompson 1990; Page 1991; Edwards and Page 1993; Stuber and Sisco 1992). Although the cost effectiveness of MAS is widely debated, the usefulness of QTL mapping for finding new favorable alleles is hard to dispute. Tanksley et al. (1989) stressed this, and illustrated how marker-assisted backcross breeding could be used to minimize linkage drag and greatly speed up the development of near-isogenic lines. Creating a picture of the distribution of favorable alleles between parents and progeny is obviously extremely useful (Young and Tanksley 1989a,b). Favorable allele frequencies

and coupling linkages of favorable alleles have been built up through years of breeding and selection and new favorable alleles for most traits are hidden in exotic germplasm. The cost of assaying marker phenotypes seems to be the limitation to using this technology more widely, and these costs are bound to decrease as marker technology advances.

One of the most important problems in plant breeding is finding new sources of favorable alleles among donor inbred lines for improving elite inbred line cultivars or the inbred parents of elite single cross hybrids (Dudley 1987; Zanoni and Dudley 1987; Gerloff and Smith 1988a,b; Stuber et al. 1993). Several statistics have been developed and have proven useful for finding new favorable alleles among donor inbred lines for developing new single-cross hybrids superior to an elite single-cross hybrid. Dudley (1984, 1987) proposed methods for estimating the number of favorable alleles in a donor inbred line which are not in either parent of an elite single-cross hybrid (n_G). This parameter and many others can be estimated from the means of the parent and donor inbred lines and their hybrids (Dudley 1987; Zanoni and Dudley 1987; Gerloff and Smith 1988a). The necessary experiments are inexpensive and useful for selecting parents for developing new inbred lines (Dudley 1987; Zanoni and Dudley 1987; Gerloff and Smith 1988a,b). So what is lacking, or rather, what is gained by mapping QTL? Parameters such as n_G summarize populations not genes. Restrictive assumptions about the genetics are needed to estimate n_G , e.g., the effects of individual genes must be equal, the effects of the genes must be completely dominant, no epistasis or linkage, and so on. Despite this, these methods seem to work extremely well because none of the biases seem to greatly affect the conclusions (Zanoni and Dudley 1987; Misevic 1989). What QTL mapping adds, however, is knowledge about the distribution of favorable alleles between parent and donor inbred lines and their progeny, estimates of gene effects without restrictive assumptions about their genetics, and marker loci linked to the genes to be selected. Once new alleles are found, they must be introgressed from donor to elite inbred lines. To develop superior single-cross hybrids, new lines must be developed which are fixed for favorable alleles in one of the parents of the elite hybrid and for new favorable alleles in the donor inbred line. MAS can be used to introgress the new favorable alleles through backcross breeding (Tanksley et al. 1989) or through pedigree or other variants of inbred line breeding methods (Lande and Thompson 1990; Lande 1992).

The usefulness of a line as a donor of new alleles cannot necessarily be determined by its phenotype alone regardless of the cultivar objective. Historically, plant breeders have had to rely only on phenotype parameters, e.g., means and variances, and experience, but these alone do not guarantee the selection of sources of favorable alleles different from those already fixed in an elite line (Stuber and Sisco 1992). This is partly why exotic germplasm has not been widely used and the gap between 'elite and exotic' germplasm constantly widens (Goodman 1986; Troyer et al. 1989; Troyer 1990).

A straightforward theoretical example illustrates the basic problem. Suppose alleles at six quantitative trait loci (A, B, C, D, E, and F) are dispersed among

Table 11 Hypothetical favorable (+) and unfavorable (-) allele distributions among inbred lines for one trait and set of target environments where the genotype mean is increased by 50 units for each favorable allele and the mean of the line homozygous for unfavorable alleles at each locus is 400 units

Locus	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆
A	++	--	++	--	--	--
B	++	--	++	--	--	--
C	++	++	++	++	--	--
D	++	++	--	--	--	--
E	++	++	++	--	--	--
F	--	++	--	--	++	--
Genotype mean	900	800	800	500	500	400

several inbred lines for some trait and an elite inbred line (P₁) is homozygous for favorable alleles at five (A, B, C, D, and E) of the six loci (Table 11). A source of favorable alleles for locus F is needed to develop a line superior to P₁ (Table 11). Two of the donor lines (P₂ and P₅) are sources of favorable alleles at the F locus; however, using genotype means alone, you could not determine the usefulness of P₅ or distinguish between P₂ and P₃, which have equal genotype means. Although the means of P₂ and P₃ are equal, a line superior to P₁ cannot be developed from the P₁ × P₃ mating, whereas a line superior to P₁ can be developed from the P₁ × P₂ and the P₁ × P₅ matings. The P₁ × P₅ mating obviously poses more of a problem for developing a line superior to P₁ than the P₁ × P₂ mating. P₂ is superior to P₅ as a donor of F locus alleles because it is fixed for favorable alleles at three additional loci (C, D, and E) and P₅ is fixed for unfavorable alleles at every other locus. But how could P₅ ever be found to be a source of useful alleles without QTL mapping? Is much of the diversity sought by plant breeders hidden within exotic germplasm as exemplified by P₅?

Transgressive segregates for new favorable alleles are often obscured when the heritability of the trait is low, the effect of the new favorable alleles is small, and the source of new favorable alleles is fixed for unfavorable alleles at most other loci. This seems plausible because favorable alleles with large effects are the easiest to find and fix. This is partly why it becomes harder to accumulate new favorable alleles in elite germplasm.

An example of favorable alleles being obscured by genes with large effects comes from QTL mapping experiments in barley where doubled haploid lines were developed from a cross between a feed grain and malting barley (Steptoe × Morex) (Hayes et al. 1993). Steptoe is an extremely important feed grain barley cultivar with very poor malting quality characteristics; however, Steptoe alleles were found to be favorable at two loci affecting malting quality. The effects of these QTL were small compared to effects of most of the other QTL, all of which were fixed for favorable alleles in Morex. Morex, of course, was

fixed for favorable alleles at more loci affecting malting quality, but Steptoe was nevertheless found to be a source of favorable alleles at two QTL, and it could conceivably be used to develop a malting barley slightly superior to Morex (Hayes et al. 1993). The problem, of course, is to introgress those two alleles, while retaining the favorable alleles fixed in Morex. Favorable alleles for many important quantitative traits could be dispersed throughout the exotic germplasm of many of our crop species, as has been shown in tomato (Paterson et al. 1992) and many other species. QTL mapping is certainly no panacea, but it gives plant breeders an objective and powerful vehicle for finding sources of new alleles and selecting the most outstanding sources of these alleles.

The multilocus questions briefly reviewed above and addressed somewhat more fully elsewhere (Knapp et al. 1993; Dudley 1993) are of critical importance for finding genes with small effects and maximizing the efficiency of marker-assisted index selection (MAS). The response to MAS is a function of the heritability of the trait and the percentage of the additive genetic variance explained by marker (σ_M^2/σ_A^2) or QTL (σ_Q^2/σ_A^2) genotypes (Smith 1967; Lande and Thompson 1990). Maximizing these is the goal of a QTL mapping experiment.

Lande and Thompson (1990) and Lande (1992) developed indexes for selection using phenotype and marker scores. Index scores are estimated by $I = b_Y Y + b_M m$ where y is the phenotype score, m is the marker score (sum of the additive effects of the QTL or marker loci), b_Y is the weight for phenotype scores, and b_M is the weight for marker scores. The weights are estimated by

$$b = P^{-1}Gd = \begin{bmatrix} \sigma_P^2 & \sigma_M^2 \\ \sigma_M^2 & \sigma_M^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_A^2 & \sigma_M^2 \\ \sigma_M^2 & \sigma_M^2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} b_Y \\ b_M \end{bmatrix} = \begin{bmatrix} \frac{\sigma_A^2 - \sigma_M^2}{\sigma_P^2 - \sigma_M^2} \\ \frac{\sigma_P^2 - \sigma_G^2}{\sigma_P^2 - \sigma_M^2} \end{bmatrix}$$

where P is the phenotypic variance-covariance matrix, G is the genotypic variance-covariance matrix, d is the vector of economic weights, σ_P^2 is the phenotypic variance, σ_A^2 is the additive genetic variance, and σ_M^2 is the additive genetic variance associated with marker loci (Lande and Thompson 1990). The efficiency of MAS relative to conventional selection is

$$\begin{aligned} \frac{R_M}{R_P} &= \frac{\sigma_P \sqrt{\sigma_A^4 - 2\sigma_A^2\sigma_M^2 + \sigma_P^2\sigma_M^2}}{\sigma_A^2 \sqrt{\sigma_P^2 - \sigma_M^2}} = \sqrt{\frac{\sigma_M^2/\sigma_A^2}{\sigma_A^2/\sigma_P^2} + \frac{(1 - \sigma_M^2/\sigma_A^2)^2}{1 - (\sigma_A^2/\sigma_P^2)(\sigma_M^2/\sigma_A^2)}} \\ &= \sqrt{\frac{\sigma_M^2/\sigma_A^2}{H} + \frac{(1 - \sigma_M^2/\sigma_A^2)^2}{1 - H(\sigma_M^2/\sigma_A^2)}} \end{aligned}$$

where i is the selection intensity, H is the heritability, the gain from selection using trait phenotypes alone is

$$R_P = i \frac{\sigma_A^2}{\sigma_P^2} \sigma_P = i \frac{\sigma_A^2}{\sigma_P} = iH\sigma_P,$$

and the gain from using a MAS index is

$$R_M = i \sqrt{d^T G b} = i \sqrt{[1 \ 0] \begin{bmatrix} \sigma_A^2 & \sigma_M^2 \\ \sigma_M^2 & \sigma_M^2 \end{bmatrix} \begin{bmatrix} \frac{\sigma_A^2 - \sigma_M^2}{\sigma_P^2 - \sigma_M^2} \\ \frac{\sigma_P^2 - \sigma_A^2}{\sigma_P^2 - \sigma_M^2} \end{bmatrix}} = i \sqrt{\frac{\sigma_A^2(\sigma_A^2 - \sigma_M^2)}{\sigma_P^2 - \sigma_M^2} + \frac{\sigma_M^2(\sigma_P^2 - \sigma_A^2)}{\sigma_P^2 - \sigma_M^2}}$$

(Lande and Thompson 1990).

The efficiency of MAS relative to conventional selection increases as σ_M^2/σ_A^2 increases and is maximum when $\sigma_M^2/\sigma_A^2 = 1.0$. As stated earlier, it should be feasible to find marker-brackets which explain 100% of the genetic variance between lines when the genome is completely covered – not necessarily saturated – by markers. Complete coverage is achieved when every segregating QTL is bracketed by markers. Then the genetic variance between marker genotypes should be equal to the genetic variance between lines.

Suppose QTL are found to lie between ten independent marker-brackets among 150 doubled haploid lines. There are 4 genotypes for each marker bracket and $4^{10} = 1,048,576$ genotypes, so the genotype of every line is almost always going to be unique unless the pairs of markers are closely linked, then some of the genotypes might be duplicated. If the marker genotype of each line is unique, the genetic variance among the marker genotypes has to be equal to the genetic variance between lines. The response to MAS is maximized by maximizing σ_M^2/σ_A^2 . This can be achieved in any experiment by randomly selecting markers to absorb the genetic variance between lines. If the selected markers are not linked to genes underlying the trait, then MAS obviously has no effect, even though $\sigma_M^2/\sigma_A^2 = 1.0$ is achieved. The markers selected obviously must be linked to bona fide QTL. The significance of an effect, e.g., a significant likelihood odds or F-statistic, does not always ensure this because of Type I and II errors (false positives and negatives).

How a genome is searched and how multiple QTL are mapped affects the implementation of MAS and the probabilities of Type I and II errors (Knapp et al. 1993). The efficiency of MAS increases as heritability decreases, while the probability of finding a given QTL decreases as heritability decreases for a given sample size. So maximizing σ_M^2/σ_A^2 is theoretically harder as heritability decreases, which is when it is most critical to do so.

The probability of encountering missing cells increases as the number of marker-brackets modelled increases. Unless there are no interactions between QTL, inferences about their effects cannot be made for many multilocus models

because of missing cells. This means the marker scores and σ_M^2 for MAS cannot be estimated from the multilocus model either. The compromise is to estimate the effects of minor QTL along with the effects of major QTL, because the sampling bias, if there is any, is much greater for the latter (Knapp et al. 1993).

7. QTL genotype by environment interactions

Many methods have been developed for characterizing line or cultivar by environment interactions (Baker 1988; Becker and Leon 1988; Freeman 1973; Lin et al. 1986; Rosielle and Hamblin 1981; Westcott 1986; Zobel et al. 1988). The goal of some of these methods is to select cultivars or lines which maximize selection gains for a set of target environments. The breeding of any trait, whether using MAS or not, entails optimizing selection for a set of *target environments* (Comstock and Moll 1963; Gardner 1963). A set of target environments is usually defined by locations and years but might encompass more definitive fixed factors, e.g., irrigation or fertilization rates.

Genotype means across *test environments* are used to select lines, populations, hybrids, and cultivars for target environments and marker and QTL alleles for MAS across target environments. Test environments – samples of years, locations, and other factors – are selected to maximize the speed of a selection cycle while minimizing the cost of testing and maximizing selection gains for target environments. Genotype by environment interactions can cause test environments to fail to maximize selection gains for target environments, with equivalent consequences for selection with and without markers. At the extreme, this happens when differences between genotypes are observed across test environments, and there are no differences across target environments. Or when differences between genotypes are not observed across test environments, and there are differences across target environments. The consequences are either to fix unfavorable alleles, or, for MAS only, to fix alleles at QTL which have no mean effect across target environments, but which had an effect across the sample of test environments used. The root of the problem with genotype by environment interactions are differences between test and target environments. Nothing can be done about the outcome of selection if test and target environments are fixed.

The nature of line by environment interactions can be explained as a function of individual QTL genotype by environment interactions. If the ranks of means of genotypes change across environments, then those genotypes manifest *crossover* genotype by environment interactions. Conversely, if differences between means of genotypes change across environments, but their ranks are constant, then those genotypes manifest *non-crossover* genotype by environment interactions. Baker (1988) proposed statistical methods (Azzalini and Cox 1984; Gail and Simon 1985) for differentiating between crossover and non-crossover interactions when testing lines or cultivars. Differentiating between crossover and non-crossover genotype by environment interactions is useful because only

crossover interactions affect how cultivars are bred and deployed (Baker, 1988); however, the outcome of selection can only be changed by redefining or reselecting test or target environments when faced with cross-over or non-crossover interactions. Non-crossover interactions do not lead to fixing unfavorable alleles – at worst QTL might be selected which do not affect the trait across target environments or QTL might be missed which affect the trait across target environments. When non-crossover interactions arise, the difference between QTL genotype means across environments is usually less than within some environments.

Lines, populations, hybrids, and cultivars often manifest genotype by environment interactions, especially when tested across diverse environments. If a significant line by environment interaction is observed for a trait, then one or more of the genes underlying the trait might manifest QTL genotype by environment interactions (Tables 1, 3, 7, and 8). The consequences of genotype by environment interactions for QTL are no different than for lines. Sorting out questions about genotype by environment interactions is obviously one of the objectives of multiple environment QTL mapping experiments. The methods described earlier can be used to sort out many of the relevant questions. Is there evidence for QTL genotype by environment interactions? What is the nature of a significant QTL genotype by environment interaction? Is there a subset of environments for which a QTL exhibits no QTL genotype by environment interaction?

Additional methods are needed to determine the nature of QTL genotype by environment interactions. These methods are hardly necessary for practicing MAS. Only the means of QTL genotypes across test environments need to be estimated to select QTL for MAS. This only fails if every QTL manifests crossover interaction and the test environments did not uncover these interactions, both of which are very unlikely for carefully selected test and target environments. Suppose, for example, a QTL manifests a crossover interaction which is not observed among the QTL genotype by test environment means, but which leads to no overall effect of the QTL within target environments, then putting selection pressure against this QTL is equivalent to selecting a neutral locus. This diminishes selection response by decreasing selection intensity, as do many errors (Edwards and Page 1993; Page 1991).

Differentiating between non-crossover and crossover QTL genotype by environment interactions might be important for optimizing MAS. Crossover QTL genotype by environment interactions could affect the outcome of MAS, whereas non-crossover interactions should be of no consequence to the outcome of MAS. Understanding and characterizing the nature of QTL genotype by environment interactions is useful for optimizing MAS or conventional selection.

Test statistics have been developed to gain insights into the nature of line by environment interactions (Baker 1988), and these can be used to gain insights into the nature of QTL genotype by environment interactions. A crossover QTL genotype by environment interaction exists when, for two QTL genotypes in

two environments, the inequality

$$\mu_{qq1} - \mu_{QQ1} > 0, \mu_{qq2} - \mu_{QQ2} < 0$$

or

$$\mu_{qq1} - \mu_{QQ1} < 0, \mu_{qq2} - \mu_{QQ2} > 0$$

are satisfied (Azzalini and Cox, 1984) where μ_{QQi} and μ_{qqi} are means of the QQ and qq genotypes in the i th environment. Azzalini and Cox (1984) developed a statistic for testing the null hypothesis of no crossover interaction. The null hypothesis of no crossover interaction for two QTL genotypes tested in two environments is rejected when the inequalities

$$\hat{\mu}_{qq1} - \hat{\mu}_{QQ1} \geq t_\alpha \hat{\sigma}_M, \hat{\mu}_{qq2} - \hat{\mu}_{QQ2} \leq -t_\alpha \hat{\sigma}_M$$

or

$$\hat{\mu}_{qq1} - \hat{\mu}_{QQ1} \leq t_\alpha \hat{\sigma}_M, \hat{\mu}_{qq2} - \hat{\mu}_{QQ2} \geq -t_\alpha \hat{\sigma}_M$$

are satisfied where

$$t_\alpha = \Phi\left\{ \left[\frac{(-2)\log(1-\alpha)}{q(q-1)e(e-1)} \right]^{1/2} \right\} = \Phi(x),$$

α is the probability of a Type I error, $\Phi(x)$ is the standard normal cumulative distribution function, $\hat{\sigma}_M^2$ is the standard error of a difference between marker genotype by environment means, q is the number of QTL genotypes, and e is the number of environments. These inequalities were defined by Azzalini and Cox (1984) as a *quadruple*. Every quadruple is then tested for a set of genotypes and environments. The null hypothesis of no crossover interaction is rejected when the inequalities are satisfied. An expedient way to implement this test is to test the quadruple for the two QTL genotypes with the largest mean difference with different signs. If either inequality is satisfied, then the null hypothesis is rejected.

There was a significant QTL genotype by environment interaction for the *Plc-iABI151* marker-bracket (Table 9 and Fig. 2). The additive effects of the QTL genotypes for Corvallis and Pullman were $\hat{\mu}_{QQ} - \hat{\mu}_{qq} = 336.8$ and $\hat{\mu}_{QQ} - \hat{\mu}_{qq} - \hat{\mu}_{qq} = -88.9$. The ranks of the QQ and qq genotypes changed across environments, and these changes were statistically significant (Table 9 and Fig. 2). The standard error of a marker by location mean was $\hat{\sigma}_M = 113.8$,

$$t_\alpha = t_{0.01} = \Phi\left\{ \left[\frac{(-2)\log(1-0.01)}{2(2-1)2(2-1)} \right]^{1/2} \right\} = \Phi(0.0005) = 2.57,$$

and $t_{0.01} \hat{\sigma}_M^2 = 2.57 (113.8) = 292.5$. The quadruple for testing the hypothesis of no cross-over QTL genotype by environment interaction is $336.8 \geq 292.5$ and $-88.9 > -292.5$, so we fail to reject the hypothesis of no crossover interaction because the magnitude of the difference between QQ and qq genotypes for Pullman was less than the test statistic. While the interaction between QTL

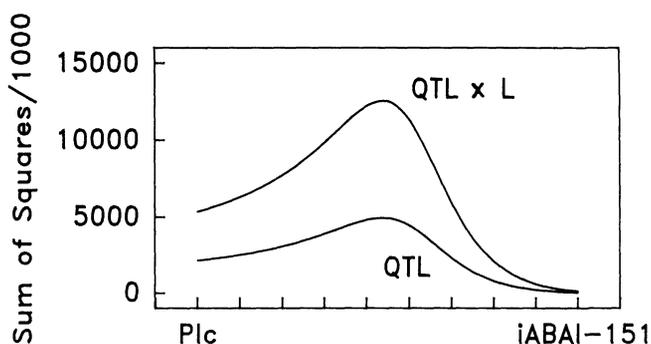


Fig 2 Seed yield (kg/ha) sum of squares for the additive (QTL) and additive by location (QTL \times L) effect of a QTL between the *Plc* and *iAB1151* locus for barley doubled haploid lines tested in Corvallis, Oregon and Pullman, Washington

genotypes and locations was significant and the estimated ranks of QTL genotype means were different at Corvallis and Pullman, the evidence does not exist to support a crossover interaction. This is consistent with what we found for individual location estimates. The QTL had no effect at Pullman, whereas it had a significant effect at Corvallis (Tables 3 and 4 and Fig. 1).

The number of quadruples which must be tested increases as the number of environments increases. The picture can become quite muddy. Some of the interactions between QTL genotypes and environments could be crossover and non-crossover for any locus or set of environments. This test (Azzalini and Cox 1984) at least shows how the nature of QTL genotype by environment interactions can be investigated.

8. Acknowledgements

This work was funded by USDA grants (#90-34213-5190, #91-37300-6529, and #92-37300-7652) and an Alexander von Humbolt Fellowship. I am grateful to Prof. Dr. Jeff Schell, Dr. Reinhard Toepfer, and Dr. Norbert Martin at the Max Planck Institute, Cologne, Germany where some of this work was done, and to Dr. William Beavis (Pioneer Hi-Bred International), Dr. Robert Stucker (University of Minnesota), and Dr. Chris Haley (University of Edinburgh) for critiquing the original manuscript.

9. References

- Azzalini, A and Cox, D R (1984) Two new tests associated with analysis of variance J R Statist Soc B 46 335-343
 Baker, R J (1988a) Tests for crossover genotype-environment interactions Can J Plant Sci 68 405-410

- Baker, R J (1988b) Analysis of genotype-environment interactions in crops *ISI Atlas Sci Animal and Plant Sci* 1 1-4
- Becker, H C and Leon, J (1988) Stability analysis in plant breeding *Plant Breed* 101 1-23
- Bulmer, M G (1980) *The Mathematical Theory of Quantitative Genetics* Oxford University Press, Oxford
- Carbonell, E A , Gerig, T M , Balansard, E and Asins, M J (1992) Interval mapping in the analysis of nonadditive quantitative trait loci *Biometrics* 48 305-315
- Carbonell, E A , Asins, M J , Baselga, M , Balansard, E and Gerig, T M (1993) Power studies in the estimation of genetic parameters and the localization of quantitative trait loci for backcross and doubled haploid populations *Theor Appl Genet* 86 411-416
- Comstock, R E and Moll, R H (1963) Genotype-environment interactions In W D Hanson and H F Robinson (eds), *Statistical Genetics and Plant Breeding NAS-NRC Publ* 982
- Doebley, J and Stec, A (1991) Genetic analysis of the morphological differences between maize and teosinte *Genetics* 129 285-295
- Doebley, J , Stec, A , Wendel, J and Edwards, M (1990) Genetic and morphological analysis of a maize-teosinte F₂ population implications for the origin of maize *Proc Natl Acad Sci U S A* 87 9888-9892
- Dudley, J W (1984) A method of identifying lines for use in improving parents of a single-cross *Crop Sci* 24 355-357
- Dudley, J W (1987) Modification of methods for identifying inbred lines useful for improving parents of elite single crosses *Crop Sci* 27 944-947
- Dudley, J W (1993) Molecular markers in plant improvement manipulation of genes affecting quantitative traits *Crop Sci* 33 660-668
- Edwards, M D and Page, N J (1993) Evaluation of marker-assisted selection through computer simulation *Theor Appl Genet* (in press)
- Edwards, M D , Stuber, C W and Wendel, J F (1987) Molecular marker facilitated investigations of quantitative trait loci in maize I Number, genomic distribution, and type of gene action *Genetics* 116 113-125
- Falconer, D S (1981) *Introduction to Quantitative Genetics* Longman, New York
- Freeman, G H (1973) Statistical methods for the analysis of genotype-environment interactions *Heredity* 31 339-354
- Freund, R J , Littell, R C and Spector, P C (1986) *SAS® System for Linear Models* SAS, Cary, North Carolina
- Gail, M and Simon, R (1985) Testing for qualitative interaction between treatment effects and patient subsets *Biometrics* 41 361-372
- Gardner, C O (1963) Estimates of genetic parameters in cross-fertilizing plants and their implications in plant breeding In W D Hanson and H F Robinson (eds), *Statistical Genetics and Plant Breeding NAS-NRC Publ* 982
- Gerloff, J and Smith, O S (1988) Choice of method for identifying germplasm with superior alleles I Theoretical results *Theor Appl Genet* 76 209-216
- Gerloff, J and Smith, O S (1988) Choice of method for identifying germplasm with superior alleles I Computer simulation results *Theor Appl Genet* 76 217-227
- Haley, C S and Knott, S A (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers *Heredity* 69 315-324
- Hanson, W D (1963) Heritability In W D Hanson and H F Robinson (eds), *Statistical Genetics and Plant Breeding NAS-NRC Publ* 982
- Hayes, P M , Liu, B -H , Knapp, S J , Chen, F , Jones, B , Blake, T , Franckowiak, J , Rasmusson, D , Sorrells, M , Ullrich, S E , Wesenberg, D and Kleinhofs, A (1993) Quantitative trait locus effects and environmental interaction in a sample of North American barley germplasm *Theor Appl Genet* 87 392-401
- Helentjaris, T , Slocum, M , Wright, S , Schaefer, A and Nienhuis, J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms *Theor Appl Genet* 72 761-769

- Jansen, R (1992) A general mixture model for mapping quantitative trait loci by using molecular markers *Theor Appl Genet* 85 252–260
- Jansen, R (1993a) Interval mapping multiple quantitative trait loci *Genetics* 135 205–211
- Jansen, R (1993b) Maximum likelihood in a generalized linear finite mixture model by using the EM algorithm *Biometrics* (in press)
- Jensen, J (1989) Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker gene loci *Theor Appl Genet* 78 613–618
- Kleinhofs, A, Kilian, A, Saghai Maroof, M A, Biyashev, R M, Hayes, P M, Chen, F Q, Lapitan, N, Fenwick, A, Blake, T K, Kanazin, V, Ananiev, E, Dahleen, L, Kudrna, D, Bollinger, J, Knapp, S J, Liu, B-H, Sorrells, M, Heun, M, Fanckowiak, J D, Hoffman, D, Skadsen, R and Steffenson, B J (1993) A molecular, isozyme, and morphological map of the barley (*Hordeum vulgare*) genome *Theor Appl Genet* 86 705–712
- Knapp, S J (1989) Quasi-Mendelian genetic analysis of quantitative trait loci using molecular marker linkage maps In G Roebelen, (ed), Proc XIIth Eucarpia Congress (European Plant Breeding Congress), Göttingen, Germany
- Knapp, S J (1991) Using molecular markers to map multiple quantitative trait loci Models for backcross, recombinant inbred, and doubled haploid progeny *Theor Appl Genet* 81 333–338
- Knapp, S J and Bridges, W C (1990) Using molecular markers to estimate quantitative trait locus parameters Power and genetic variances for unreplicated and replicated progeny *Genetics* 126 769–777
- Knapp, S J, Bridges, W C and Birkes, D (1990) Mapping quantitative trait loci using molecular marker linkage maps *Theor Appl Genet* 79 583–592
- Knapp, S J, Bridges, W C and Liu, B-H (1992) Mapping quantitative trait loci using nonsimultaneous and simultaneous estimators and hypothesis tests In J S Beckmann and T S Osborn (eds), *Plant Genomes Methods for Genetic and Physical Mapping*, pp 209–237 Kluwer Academic Publishers, Dordrecht, The Netherlands
- Knott, S A and Haley, C S (1992) Aspects of maximum likelihood methods for the mapping of quantitative trait loci in line crosses *Genet Res* 60 139–151
- Lande, R (1992) Marker-assisted selection in relation to traditional methods of plant breeding In H T Stalker and J P Murphy (eds), *Plant Breeding in the 1990s Proc Symp Plant Breeding in the 1990s*, C A B International, Wallingford, U K
- Lande, R and Thompson, R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits *Genetics* 124 743–756
- Lander, E S and Botstein, D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps *Genetics* 121 185–199
- Landry, B S, Kesseli, R V, Fararra, B and Michelmore, R W (1987) A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphisms, isozymes, disease resistance, and morphological markers *Genetics* 116 331–337
- Lin, C S, Binns, M R and Lefkovich, L P (1986) Stability analysis Where do we stand? *Crop Sci* 26 894–900
- Lou, Z W and Kearsy, M J (1989) Maximum likelihood estimation of linkage between a marker gene and a quantitative trait locus *Heredity* 66 117–124
- Martinez, O and Curnow, R N (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers *Theor Appl Genet* 85 480–488
- McLachlan, G J and Basford, K E (1988) *Mixture Models Inference and Applications to Clustering* Marcel Dekker, New York
- Misevic, D (1988) Evaluation of three test statistics used to identify maize inbred lines with new favorable alleles not present in elite single crosses *Theor Appl Genet* 77 402–408
- Nienhuis, J, Helentjaris, T, Slocum, M, Ruggero, B and Schaefer, A (1987) Restriction fragment length polymorphism analysis of loci associated with insect resistance in tomato *Crop Sci* 27 797–803
- Ooijen, J W (1992) Accuracy of mapping quantitative trait loci in autogamous species *Theor Appl Genet* 84 803–811
- Osborn, T C, Alexander, D C and Fobes, F (1987) Identification of restriction fragment length

- polymorphisms linked to genes controlling soluble solids content in tomato fruit *Theor Appl Genet* 73 350–356
- Page, N P (1991) A computer simulation evaluation of the utility of marker-assisted selection Ph D thesis, University of Minnesota, St Paul, MN
- Paterson, A H , Lander, E S , Hewitt, J D , Peterson, S , Lincoln, S E and Tanksley, S D (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms *Nature* 335 721–726
- Paterson, A H , Damon, S , Hewitt, J D , Zamir, D , Rabinowitch, H D , Lincoln, S E , Lander, E S and Tanksley, S D (1991) Mendelian factors underlying quantitative traits in tomato comparison across species, generations, and environments *Genetics* 127 181–197
- Rodolphe, F and Lefort, M (1993) A multi-marker model for detecting chromosomal segments displaying QTL activity *Genetics* 134 1277–1276
- Rosielle, A A and Hamblin, J (1981) Theoretical aspects of selection for yield in stress and non-stress environments *Crop Sci* 21 943–946
- Satterthwaite (1946) An approximate distribution of estimates of variance components *Biometrics Bull* 2 110–114
- Searle, S (1971) *Linear Models* John Wiley and Sons, New York
- Shukla, G K (1972) Some statistical aspects of partitioning genotype-environmental components of variability *Heredity* 29 237–245
- Simpson, S P (1989) Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines *Theor Appl Genet* 77 815–819
- Smith, C (1967) Improvement of metric traits through specific genetic loci *Anim Prod* 9 349–358
- Stuber, C W and Sisco, P H (1991) Marker-facilitated transfer of QTL alleles between elite inbred lines and responses in hybrids *Ann Corn Sorghum Res Conf* 46 104–113
- Stuber, C W , Edwards, M D and Wendel, J F (1987) Molecular-marker facilitated investigations of quantitative trait loci in maize II Factors influencing yield and its component traits *Crop Sci* 27 639–648
- Stuber, C W , Lincoln, S E , Wolff, D W , Helentjaris, T and Lander, E S (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers *Genetics* 132 823–839
- Soller, M and Beckmann, J S (1990) Marker-based mapping of quantitative trait loci using replicated progenies *Theor Appl Genet* 80 205–208
- Titterton, D M , Smith, A F M and Makov, U E (1985) *Statistical Analysis of Finite Mixture Distributions* Wiley and Sons, New York
- Troyer, A F (1990) A retrospective view of corn genetic resources *J Hered* 81 17–24
- Troyer, A F , Openshaw, S J and Knittle, K H (1988) Measurement of genetic diversity among popular commercial corn hybrids *Crop Sci* 28 481–485
- van Ooijen, J W (1992) Accuracy of mapping quantitative trait loci in autogamous species *Theor Appl Genet* 84 803–811
- Weller, J I (1986) Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers *Biometrics* 42 627–640
- Westcott, B (1986) Some methods of analysing genotype-environment interaction *Heredity* 56 243–253
- Young, N D and Tanksley, S D (1989a) Restriction fragment length polymorphism maps and the concept of graphical genotypes *Theor Appl Genet* 77 95–101
- Young, N D and Tanksley, S D (1989b) RFLP analysis of the size of chromosomal segments retained around the Tm-2 locus of tomato during backcrossing *Theor Appl Genet* 77 353–359
- Zanoni, U and Dudley, J W (1989) Comparison of different methods of identifying inbreds useful for improving elite maize hybrids *Crop Sci* 29 577–582
- Zhang, W and Smith, C (1992) Computer simulation of marker-assisted selection utilizing linkage disequilibrium *Theor Appl Genet* 83 813–820
- Zobel, R W , Wright, M J and Gauch, H G Jr (1988) Statistical analysis of a yield trial *Crop Sci* 80 388–393

5. Breeding multigenic traits

CHARLES W. STUBER

U S Department of Agriculture, Agricultural Research Service, Department of Genetics, North Carolina State University, Raleigh, NC 27695-7614, U S A

Contents

1 Introduction	97	6 Heterosis and marker-facilitated enhancement of heterosis	106
2 Challenges facing plant breeders	98	7 Marker-facilitated introgression (backcrossing)	111
3 Hybrid predictions	99	8 Conclusions	112
4 Marker allele frequency changes	102	9 References	113
5 Quantitative trait manipulation in F ₂ populations	103		

1. Introduction

Increases in the productivity of food, feed, and fiber in domesticated crop plants can be attributed to collective plant breeding efforts over many millennia. These increases (many have been dramatic) have resulted from artificial selection, either conscious or unconscious, on the phenotypes of the targeted species. Until the 20th century, plant breeding was largely an art with little or no knowledge of genetic principles. Although plant improvement since the rediscovery of Mendel's principles has involved both art and science, the contributions of science will undoubtedly assume a much greater role as new technology becomes more widely used and as additional gains in agricultural productivity are required at an ever-increasing pace.

If we consider the increase in maize grain yield production in the United States, for example, we note that the gain for the time span from 1930 to 1980 averaged nearly one q ha⁻¹ yr⁻¹ with about three-fourths of that gain attributed to genetic improvement (Duvick 1984; Russell 1984). The remainder is attributed to changes in cultural practices such as increased rates of mineral fertilizers and the use of herbicides for weed control and pesticides for control of insects and diseases. Because there are limited opportunities for gains resulting from changes in cultural practices, future gains in the productivity of most crops may depend almost entirely on genetic improvements. In fact, environmental concerns may cause a reduction in the use of agricultural

chemicals and fertilizers. Therefore, plant breeders will need to develop and apply new technology (such as DNA-based markers) at a faster pace to more effectively improve crop plants for the ever increasing global human population.

Most economically important plant traits, such as grain or forage yield, are classified as multigenic or quantitative. Even traits considered to be more simply inherited, such as disease resistance, may be 'semi-quantitative' for which trait expression is governed by several genes (e.g., a major gene plus several modifiers). The challenge to strategically use new technology (such as DNA-based markers) to increase the contribution of 'science' to the 'art + science' equation for plant improvement therefore applies to most, if not all, traits of importance in plant breeding programs.

Historically, one of the first questions in quantitative genetics was whether the inheritance of these continuously distributed traits was Mendelian (Comstock 1978). Obviously, the answer to this question has major implications in the consideration of the use of markers for plant breeding programs. Evidence reported by East and Hayes (1911), Emerson and East (1913), and others contributed to the rejection of the 'blending' inheritance hypothesis and to the conclusion that Mendelian principles apply to quantitative as well as to qualitative traits. Over the past 80 years, both plant and animal geneticists have obtained convincing evidence for the shaping of the general model that embraces the multiple-factor hypothesis for quantitative traits (with genes located in chromosomes and hence sometimes linked, and incomplete heritability because of the contribution of environmental factors to total phenotypic variation).

2. Challenges facing plant breeders

Plant breeders are faced with numerous challenges in their improvement programs, many of which might be met with the development and application of new technology. Some of these challenges that have stimulated research in the application of marker technology relate to:

1. prediction of hybrid performance,
2. identification of useful genetic factors in divergent populations or lines (such as exotic accessions),
3. introgression of desired genetic factors into elite breeding lines,
4. improvement of recurrent selection programs based on phenotypic responses.

In order to meet these challenges, the primary emphasis has been on the development of new tools, such as DNA-based markers, with the major focus on the improvement of breeding precision and efficiency.

In spite of the fact that numerous investigations have been conducted on the inheritance of multigenic traits (using primarily classical biometrical methods), plant breeders typically have little information on: (1) the number of genetic

factors (loci) influencing the expression of the traits, (2) the chromosomal location of these loci, (3) the relative size of the contribution of individual loci to trait expression, (4) pleiotropic effects, (5) epistatic interactions among genetic factors, and (6) variation of expression of individual factors in different environments. In order to achieve the maximum benefit from marker-based procedures for the manipulation and improvement of multigenic traits, an increased understanding of the genetic bases underlying quantitative trait variation will be necessary.

Marker-based technology already is providing scientists with a powerful approach for identifying and mapping quantitative trait loci (QTLs) and should ultimately lead to a better understanding of genetic phenomena such as epistasis, pleiotropy, and heterosis. A number of recent investigations, particularly in maize and tomato, have provided some clues to understanding such phenomena (e.g., Edwards et al. 1987; Stuber et al. 1987; Paterson et al. 1990, 1991; Abler et al. 1991; Koester 1992; Edwards et al. 1992; Stuber et al. 1992). It must be acknowledged, however, that studies such as these have identified and mapped only rather large chromosomal segments (in most cases probably 20 to 30 cM long). Although results from such studies may be adequate for many plant breeding endeavors, novel approaches will be necessary to identify individual genes and quantify individual gene action and interactions among genes.

3. Hybrid predictions

Heterosis (or hybrid vigor) is a major reason for the success of the commercial maize industry as well as for the success of breeding efforts in many other crop and horticultural plants. Development of inbred lines suitable for use in production of superior hybrids is very costly and requires many years in traditional breeding programs. Much of the developmental effort is devoted to field testing of newly created lines in various single-cross combinations to identify those lines with superior combining ability.

The search for a reliable method for predicting hybrid performance without generating and testing hundreds or thousands of single-cross combinations has been the goal of numerous marker studies, particularly in maize. For example, in several earlier studies, correlations between isozyme allelic diversity and grain yield were estimated in single-cross hybrids derived from commercially used lines (Hunter and Kannenberg 1971; Heidrich-Sobrina and Cordeiro 1975; Gonella and Peterson 1978; Hadjinov et al. 1982). With 11 or fewer isozyme marker loci and 15 or fewer inbred parental lines, the estimated correlations between isozyme allelic diversity and specific combining ability were low and nonsignificant. Even in a much larger study in which 100 maize hybrids derived from 37 elite lines were used to evaluate associations of hybrid yield performance with allelic diversity at 31 isozyme loci, an R-square value of only 0.36 was reported by Smith and Smith (1989). Also, no association was found

between hybrid grain yield and isozyme diversity in a recent study of six enzyme marker loci in 75 F₁ rice (*Oryza sativa* L.) hybrids (Peng et al. 1988).

Frei et al. (1986a) studied 114 single-cross maize hybrids with a somewhat different orientation from those discussed above. They compared the merits of using isozyme marker diversity with those of pedigree diversity for predicting hybrid yield performances. Genotypes of 21 isozyme loci were used to classify inbred line pairs into similar and dissimilar groups. These isozyme diversity groups were further subdivided into similar and dissimilar pedigree classes based on commonality of pedigree background between line pairs. Comparisons of the isozyme diversity groups showed that average grain yield of hybrids with dissimilar isozymes was significantly higher (10% greater) than hybrids with similar isozyme genotypes. However, hybrid yields for the dissimilar pedigree class averaged about 37% more than for the hybrids in the similar pedigree class. Although isozyme dissimilarity was significantly associated with higher yield in the single-cross hybrids tested, the investigators concluded that the useful predictive value of these markers would be limited primarily to lines with similar pedigrees. In another isozyme diversity/hybrid performance study in maize, Lamkey et al. (1987) investigated F₁ hybrids among 24 high-yielding and 21 low-yielding lines from a group of 247 inbred lines derived using single-seed descent in the Iowa Stiff Stalk Synthetic (BSSS) population. Allelic differences at 11 isozyme marker loci were not predictive of hybrid performance in comparisons among crosses of high- and low-yielding lines.

The above studies suggest that isozyme genotypes provide limited value in the prediction of hybrid performance in crops such as maize and rice. Several factors may contribute to this somewhat disappointing conclusion. For example, the small number of isozyme loci assayed in most of the studies would effectively mark only a small fraction of the genome. Therefore, only a limited proportion of the genetic factors contributing to the hybrid response would be sampled. More importantly, it is unlikely that these marker loci affect the phenotypic expression of the targeted quantitative trait directly; rather they serve to identify adjacent (linked) chromosomal segments. Certainly allelic differences at marker loci do not assure allelic differences at linked QTLs. For a limited number of markers to be useful as predictors for hybrid performance, the effects of QTL 'alleles' linked to specific marker alleles must be ascertained.

Also, it should be noted that the type of gene action associated with specific QTLs will affect the predictive value of linked marker loci. In maize populations developed from crosses of two inbred lines, it has been shown that the number of heterozygous marker loci is positively correlated with grain yield of F₂ plants or backcross families (Edwards et al. 1987; Stuber et al. 1992). These results corroborated other data that implicate dominant (or even overdominant) types of gene action as the predominant contributor to the expression of grain yield in maize. In such cases, marker allele diversity that reflects linked QTL 'allele' diversity should be predictive of grain yield responses. However, for traits governed largely by additive gene action (and this

type of gene action might prevail for some loci affecting grain yield) the heterozygous QTL genotype would not be the most favorable. Again, as stated in the preceding paragraph, effective prediction of hybrid performance based on markers requires knowledge of QTLs linked to the markers.

Furthermore, it should be stressed that the level of linkage integrity must not be overlooked in the consideration of markers for hybrid predictions. For example, if the proposed hybrids are derived from lines produced from a randomly mated population, or if the lines comprise some subset of publicly available inbreds, then the associations between marker alleles and QTL alleles might be expected to be essentially random, i.e., near linkage equilibrium. For marker-based procedures to be successful for predictive or selective purposes for complexly inherited traits, such as grain yield, the genome should be well saturated with uniformly spaced markers and/or a high level of linkage disequilibrium must exist.

In the study discussed previously by Smith and Smith (1989), associations of grain yield with diversity of RFLP genotypes also were measured in the more than 100 hybrids derived from 37 elite maize inbred lines. Plots of F_1 grain yield against RFLP diversity, based on 230 marker loci, showed an R-square value of 0.87. This value presents a striking contrast between the use of 230 RFLP marker loci versus 31 isozyme loci (with an R-square value of 0.36) for the prediction of hybrid performance. However, it is important to note that even with 230 RFLP markers, yields varied from 8.15 to 10.66 t/ha (130 to 170 bushels/acre) for the subset of hybrids with the maximum detected 'distance' (0.70 to 0.80 on a scale ranging from 0.00 to 0.80) between the parental lines. Most breeders would be working with similar subsets of largely unrelated lines for which, again, marker diversity alone does not appear to be very satisfactory for predictive purposes.

In another maize study, Melchinger et al. (1990) compared RFLP genotypes at 82 marker loci with field data on 67 hybrids reported earlier by Darrah and Hallauer (1972). Twenty inbred lines were involved in the parentage of the hybrids. They concluded that associations of hybrid yield, heterosis, and specific combining ability with multilocus heterozygosity of RFLP loci generally were too weak to be useful as a supplementary tool for predicting yield performance of crosses between unrelated lines. In addition, they concluded that for unrelated lines, genetic distance measures based on a large number of RFLPs uniformly distributed throughout the genome are not markedly superior to those based on a small number of isozymes for predicting hybrid yield. Thus, their results show that better marker coverage alone will not increase predictive power substantially. Melchinger et al. further state that 'it seems necessary to employ specific markers for those segments that significantly affect the expression of heterosis for grain yield'.

Dudley et al. (1992) reported a study for which the major objectives were to evaluate methods of using molecular marker data to: (1) identify parents useful for improving a single-cross hybrid and (2) compare marker genotypic means measured at the inbred level to those measured at the hybrid level.

Genotypic data from 14 isozyme and 52 RFLP marker loci were compared with field performance data from a diallel mating design of 14 maize inbreds in their investigations. They found that marker genotypic differences measured in inbreds were positively correlated with differences measured in hybrid backgrounds. However, these correlations were only slightly higher than those between phenotypic midparent and hybrid values. Their findings suggested that genotypic differences may be useful for preliminary selection of loci and alleles for possible improvement of hybrids but probably will not accurately predict final performance of a hybrid. They concluded that the number of unique alleles in a donor line was not a good measure for identifying lines that have value for improving a single cross and stated that 'uniqueness of alleles does not necessarily indicate the presence of a favorable QTL'. Thus, results from this study corroborated results from many earlier attempts to correlate marker allele diversity of parental lines with hybrid performance.

4. Marker allele frequency changes

An earlier approach in the development of markers for plant breeding purposes was based on the association of isozyme allelic frequencies with phenotypic changes of targeted quantitative traits in long-term recurrent selection experiments. Studies in this area were designed not only to search for the presence of marker associations with quantitative traits but also to ascertain the strength of these associations. Mapping QTLs came later in development of breeding strategies based on marker technology.

Changes in allelic frequencies at a large number of isozyme marker loci were monitored over different cycles of long-term recurrent selection in several populations of maize (Stuber and Moll 1972; Stuber et al. 1980). Statistically significant changes were reported at eight marker loci which were greater than would be expected from drift acting alone. More importantly, these frequency changes were highly correlated with phenotypic changes in the selected trait, grain yield. Associations between isozyme marker genotypes and several agronomic and morphological traits in maize recurrent selection experiments also have been reported by Pollak et al. (1984), Kahler (1985), and Guse et al. (1988).

Results from studies such as these led to the hypothesis that manipulation of allelic frequencies at selected isozyme marker loci should produce significant responses in the correlated trait. An investigation to test this hypothesis was based on the results from the study reported by Stuber et al. (1980). This investigation, which was conducted in an open-pollinated maize population, showed that selections based solely on manipulations of allelic frequencies at seven isozyme loci significantly increased grain yield (Stuber et al. 1982). Ear number, a trait highly correlated with grain production, also increased based on these manipulations. Frei et al. (1986b) conducted a somewhat similar study in a population generated from a composite of elite inbred lines. They reported

that manipulations of isozyme allelic frequencies produced responses nearly equal to responses based on phenotypic selection in the same population.

Although these marker-allele manipulation studies produced statistically significant results, the findings were not dramatic. In both of the studies listed above, the level of linkage disequilibrium between marker loci and QTLs was probably low because the target populations had been subjected to several generations of random mating prior to the investigation. This undoubtedly reduced the effectiveness of the marker loci for manipulating the associated quantitative traits. Although these studies could be considered to be only mildly successful, the results were sufficiently positive to stimulate further investigations using larger numbers of markers in more structured types of populations. The impetus for further study and development of marker-technology related to the breeding of quantitative traits is probably the major contribution that can be ascribed to these earlier attempts.

5. Quantitative trait manipulation in F_2 populations

In the marker-facilitated research conducted in the maize genetics program at Raleigh, North Carolina, QTLs have been identified and mapped in 15 F_2 populations derived from seven elite inbred lines and five inbred lines with a partial exotic (Latin American, expected to be 50%) component (Edwards et al. 1987; Stuber et al. 1987; Abler et al. 1991; Edwards et al. 1992; and unpub. data). Both isozyme and RFLP marker loci were used in these studies, although the earlier studies used only isozymes. Measurements recorded on individual plants in the field experiments included dimensions, weights, and counts of numerous vegetative and reproductive plant parts as well as silking and pollen shedding dates.

Findings from these F_2 studies showed that QTLs affecting most of the quantitative traits evaluated were generally distributed throughout the genome, however, certain chromosomal regions appeared to contribute greater effects than others to trait expression. For example, major factors associated with the expression of grain yield were detected in the vicinity of *Mdh4*, *Adh1*, and *Phi1* on chromosome 1L; *Dial1* on chromosome 2S; *Mdh3* and *Pgd2* on chromosome 3L; *Amp3*, *Mdh5*, and *Pgm2* on chromosome 5S; *Idh1* on chromosome 8L; and *Acp1* on chromosome 9S. Not all of the chromosome regions have been well marked and presumably major factors also may have been segregating in regions of the genome devoid of marker loci in these studies. In two instances, nearly 1900 plants were genotyped and evaluated for more than 80 quantitative traits in each population. One of these F_2 populations was derived from the cross of CO159 with Tx303 (denoted as COTX); the other from the cross of T232 with CM37 (denoted as CMT).

Data from these F_2 populations provided a much stronger case for marker-based selection for manipulating quantitative traits than that from the earlier isozyme allelic frequency studies. To evaluate the efficacy of marker-facilitated

manipulation, data from the COTX and the CMT F₂ population studies were used (Stuber and Edwards 1986). Selections were based solely on the genotypes of the F₂ plants evaluated in the mapping studies. Evaluations of selection response were then made on progenies of these open-pollinated F₂ plants. Phenotypic (mass) selection, based solely on the phenotypic expression of each F₂ plant, also was conducted for comparison with the marker-facilitated selection responses. The evaluations of the selection responses were made at three locations in North Carolina in the year after the individual F₂ plants were studied.

In these marker-based selection studies, several traits were manipulated and a breeding value was determined for each of 15 markers for every F₂ plant (about 1900 in each population) for each trait. This breeding value was determined by calculating one-half the difference between the quantitative trait means of the two homozygous classes for the marker locus. Plants homozygous for the favorable marker allele were assigned a plus breeding value for that marker locus; those plants homozygous for the unfavorable allele were given a minus value. Marker heterozygotes received a value of zero. A composite breeding score was then calculated for each plant/trait combination by totalling the individual breeding values for the 15 isozyme marker loci. Based on these composite scores, individual plants were chosen to provide divergent selection groups (positive and negative) for each trait, which included grain yield, ear number, and ear height.

In Table 1 it can be noted that for the CMT F₂ population, the mean of the increased-yield (genotypic-positive) entry was about 40% greater than the mean for the decreased-yield (genotypic-negative) entry. Also, the mean yield of the positive entry was about 20% greater than the mean of the unselected check (a sample of the open-pollinated population from the same F₂). It can also be noted that the response to phenotypic (mass) selection was nearly identical to the genotypic response. In the COTX population, responses to genotypic selection were similar to those for CMT, but were not so striking. This was expected because considerably more variation was accounted for by the marker loci for grain yield in CMT than in COTX. Ear height and ear number are highly correlated with grain yield and the responses of these two traits to selection for yield reflected this close association (Table 1).

To further analyze the marker-assisted selection response, marker allelic frequencies were calculated in each of the selected populations. In the CMT population, frequencies of the favorable alleles at the 15 marker loci averaged 0.38 greater in the population selected for increased (genotypic-positive) yield than in the one selected for decreased (genotypic-negative) yield (Edwards and Stuber, unpub.). The divergent selection showed differences in allelic frequencies ranging from 0.02 to 0.73 for individual loci. As expected, loci showing the greatest differences were those that accounted for the largest proportion of the phenotypic variation for grain yield.

Frequency differences between the divergent populations derived from phenotypic (mass) selection averaged 0.13, about one-third the 0.38 value for

Table 1. Means for three traits following divergent genotypic (marker-locus) and phenotypic (mass) selection for grain yield in CMT (T232 × CM37) F₂ population.

Selection criterion	Trait means		
	Grain yield (g/plant)	Ear height (cm)	Ear number
Genotypic-positive	151.2	73.5	1.48
Genotypic-negative	107.7	47.1	1.20
Phenotypic-positive	151.7	68.5	1.43
Phenotypic-negative	122.4	57.8	1.28
Check-F ₂ randomly mated	127.2	59.2	1.35
S.E. \bar{d} ^a	6.4	2.2	0.04

^a Standard error of mean difference.

the genotypic selection in the CMT populations even though the overall selection responses were similar. These findings imply that loci in the unmarked regions of the genome, which would not be expected to respond in the marker-based scheme, contributed to the phenotypic selection response. Results in the COTX populations were similar to those for CMT, but magnitudes of frequency differences were less in COTX, which corresponds to the smaller response to selection as noted above.

From this investigation, it was concluded that marker-facilitated selection (based on 15 isozyme marker loci which probably represent no more than 30 to 40% of the genome) was as effective as phenotypic selection which would be expected to involve the entire genome. Furthermore, the results imply that a significant increase in the relative effectiveness of marker-based selection could be reasonably expected if the entire genome were marked with uniformly distributed loci (e. g., every 10 to 20 cM).

Although these results conclusively demonstrated that quantitative traits, such as grain yield, can be manipulated using only genotypic (marker) data, Lande and Thompson (1990) have provided the theory and shown analytically that the maximum rate of improvement may be obtained by integrating both phenotypic and marker data. In their investigation several selection indices were derived that maximize the rate of improvement in quantitative characters under several schemes of marker-assisted phenotypic selection (including the use of phenotypic data on relatives). They also analyzed statistical limitations on the efficiency of marker-assisted selection, which included the precision of the estimated associations between marker loci and QTLs as well as sampling errors in estimating weighting coefficients in the selection indices.

Findings from Lande and Thompson's (1990) investigation showed that (on a single trait) the potential selection efficiency by using a combination of molecular and phenotypic information (relative to standard methods of phenotypic selection) depends on the heritability of the trait, the proportion of additive genetic variance associated with the marker loci, and the selection

scheme. The relative efficiency of marker-assisted selection is greatest for characters with low heritability if a large fraction of the additive genetic variance is associated with the marker loci. Limitations that may affect the potential utility of marker-assisted selection in applied breeding programs include: (1) the level of linkage disequilibria in the populations which affects the number of marker loci needed, (2) sample sizes needed to detect QTLs for traits with low heritability, and (3) sampling errors in the estimation of relative weights in the selection indices. Although recent advances in molecular genetics have promised to revolutionize agricultural practices, Lande and Thompson state 'There are, however, several reasons why molecular genetics can never replace traditional methods of agricultural improvement, but instead should be integrated to obtain the maximum improvement in the economic value of domesticated populations.' Their results, however, are encouraging and support the use of DNA-based markers to achieve substantial increases in the efficiency of artificial selection.

6. Heterosis and marker-facilitated enhancement of heterosis

As stated previously, heterosis (or hybrid vigor) has been a major contributor to the success of the commercial maize industry and is an important component of the breeding strategies in many crops and horticultural plants. The term, heterosis, was coined by G. H. Shull and first proposed in 1914 (see Hayes 1952) and normally is defined in terms of F_1 hybrid superiority over some measure of the performance of one or both parents. Possible genetic explanations for this phenomenon include true overdominance (i.e., single loci for which two alleles have the property that the heterozygote is truly superior to either homozygote), pseudo-overdominance (i.e., closely linked loci at which alleles having dominant or partially dominant advantageous effects are in repulsion linkage phase), or possibly certain types of epistasis.

An extensive field and laboratory investigation (which will be referred to as the 'heterosis' study) was reported by Stuber et al. (1992) in which a major goal was to obtain data that might lay a foundation for understanding the basis of this important phenomenon in maize. In this study, 76 marker loci (nine isozyme and 67 RFLP) that represented from 90 to 95% of the genome were used to identify and map QTLs contributing to heterosis in the cross between the elite inbred lines, B73 and Mo17. Experimental materials were derived by backcrossing 264 F_3 lines (developed by single-seed descent) to each of the two parental inbreds. Phenotypic data were recorded on the backcross progenies in six diverse environments (four in North Carolina, one in Illinois, and one in Iowa). Evaluations of phenotypic effects associated with these QTLs showed that for the trait, grain yield, the heterozygote had a higher phenotype than the respective homozygote (with only one exception), suggesting not only overdominant gene action (or pseudo-overdominance), but also that these detected QTLs contributed significantly to the expression of heterosis. There

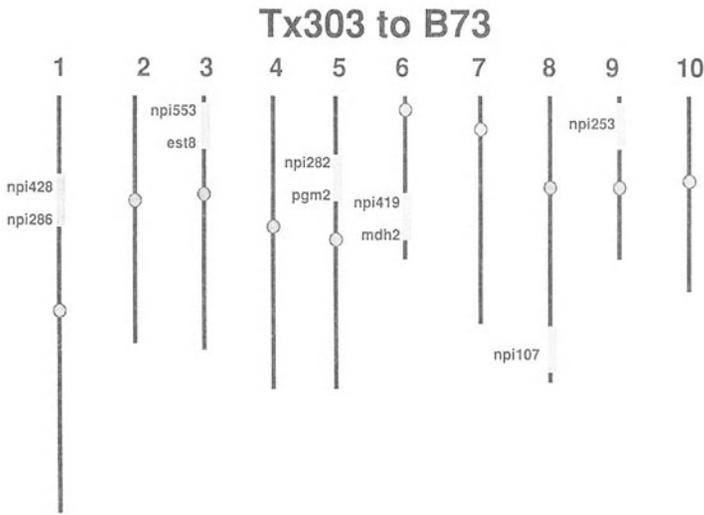


Fig. 1. Chromosomal locations of segments identified in the inbred line Tx303 for transfer into inbred line B73.

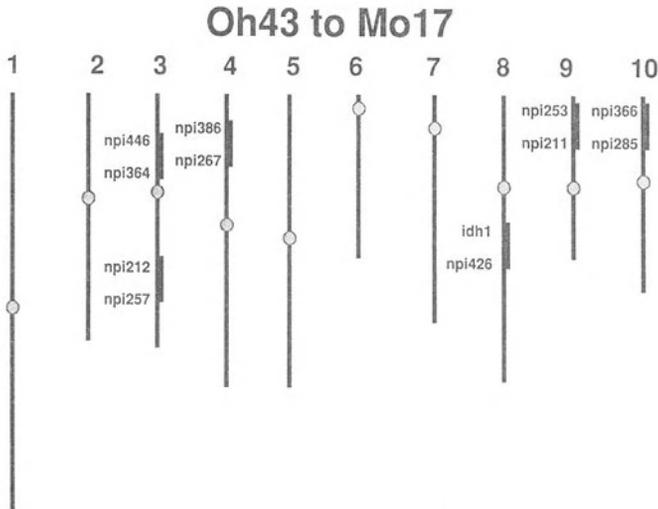


Fig. 2 Chromosomal locations of segments identified in the inbred line Oh43 for transfer into inbred line Mo17.

was little evidence for genotype (QTL) by environment interaction even though the mean yield levels for the environments sampled varied from 3.95 to 7.24 q ha⁻¹ (63 to 116 bushels per acre).

The goal of a companion study conducted with the ‘heterosis’ investigations was to use molecular markers to identify and locate genetic factors in two other

elite inbred lines, Oh43 and Tx303, that would be expected to enhance the heterotic response for grain yield in the B73 × Mo17 single-cross hybrid. Partial results from this study have been reported by Stuber and Sisco (1991). Experimental materials were developed by hybridizing the two lines, Oh43 and Tx303, and selfing for two generations to create 216 single-seed descent F₃ lines. One plant from each of these lines was testcrossed to both B73 and Mo17, and then selfed to provide progeny for marker genotyping. Phenotypic data were recorded on the testcross progenies in the same six environments used for the 'heterosis' study. By making appropriate comparisons among phenotypic means for marker classes (from testcross progenies in this study and backcross progenies in the 'heterosis' study), six chromosomal segments were identified in Tx303 and another six in Oh43 that (if transferred into B73 and Mo17, respectively) would be expected to enhance the B73 × Mo17 hybrid response for grain yield (Figs. 1 and 2).

Three backcross generations (two marker-facilitated) were used for the transfer (introgression) of identified chromosomal segments into the target lines, B73 and Mo17. Marker analyses in the BC₂ and BC₃ families were used to select individuals for the succeeding backcross and selfing generations, respectively. Individuals were selected with the desired genotype for markers bracketing the donor segment to be transferred and with the recipient marker genotype in the remainder of the genome. At least one marker was monitored on each of the 20 chromosome arms. After the third backcross, selected plants were selfed for two generations and then crossed to appropriate testers to evaluate the contribution of the introgressed segments. Marker genotyping followed similar procedures in the two backcross generations and the two selfing generations. However, if a marker locus became fixed in a line, that marker was not evaluated in that line in the succeeding generation and only segregating loci were analyzed. This reduced the laboratory analyses considerably.

After the second selfing generation, 141 BC₃S₂ 'enhanced' B73 lines were identified for testcrossing to the original Mo17. Likewise, 116 BC₃S₂ 'enhanced' Mo17 lines were identified and testcrossed to the original B73. These 257 testcross hybrids were evaluated in replicated field plots at three locations in North Carolina. Testcross data were compared with the check hybrid (B73 × Mo17) derived from the original B73 and Mo17 parental lines.

Results from these field evaluations were reported by Stuber and Sisco (1991). Numbers of testcross hybrids that yielded the same as the check or differed by one or two standard errors (of the mean difference) are shown in Figs. 3a and b. It is obvious that the results are skewed in the positive direction.

Figure 3a shows that for the 141 hybrids developed by crossing 'enhanced' B73 with normal Mo17, 45 (32%) yielded more than the check hybrid by at least one standard deviation. This can be contrasted with 15 (11%) that showed less grain yield than the check hybrid.

Results from evaluations of the 116 testcrosses derived from crossing the 'enhanced' Mo17 lines with normal B73 are summarized in Fig. 3b and are skewed even more in the positive direction. Of these 116 hybrids, 51 (44%)

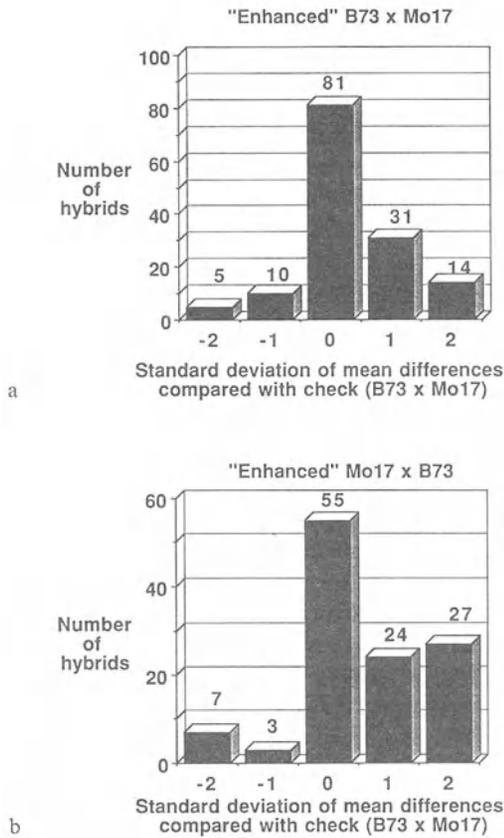


Fig. 3. (a) Bar graph comparing 'Enhanced' B73 \times normal Mo17 testcrosses with check hybrid (normal B73 \times normal Mo17); (b) Bar graph comparing 'Enhanced' Mo17 \times normal B73 testcrosses with check hybrid (normal B73 \times normal Mo17).

showed a greater grain yield than the check as contrasted with only 10 (<9%) that yielded less than the check by one or more standard deviations. Yields of eight of the highest yielding testcross hybrids using 'enhanced' Mo17 as one parent are shown in Table 2.

Although the results from this marker-facilitated introgression experiment are striking, some might have anticipated a higher success rate. It should be noted, however, that a restricted amount of resources were available for the study. Only about 600 backcross families for each of the two lines could be evaluated. Therefore, not all of the targeted segments could be transferred and most of the 'enhanced' lines have only one to three introduced segments. Further research will eventually incorporate various combinations of all of the targeted segments. Also, by following only one marker per chromosome arm to monitor the recovery of the recipient genome in the non-targeted regions, it is likely that some of the 'enhanced' lines have incorporated unidentified regions

Table 2 Grain yields of eight high yielding single-cross hybrids developed by crossing 'Enhanced' Mo17 lines to normal B73

Line	Segment (s) from Oh43	Hybrid yield bu/A
<i>Group 1</i>		
241-8-3-1-7-5	4S	117.4
241-8-3-1-7-8	4S	118.5
241-8-3-1-7-10	4S	121.0
241-8-3-1-7-13	4S	119.5
Check (B73 × Mo17)		109.1 ^b
<i>Group 2</i>		
241-112-3-3-3-7	3S, 10S	113.5
241-339-1-2-9-1	3S, 4S	118.3
241-339-1-2-9-2	3S, 4S	115.5
241-339-1-2-15-3	3S, 4S, 8L	112.4
Check (B73 × Mo17)		104.7 ^b
S E \bar{a} ^a		3.3

^a Standard error of mean difference

^b Hybrids were compared with the check in the block in which the hybrids were grown. Check means differ because of the position of the blocks in the experimental fields

from the donor lines that might include deleterious genes. In addition, the 'tightness' of the linkage of the targeted segments with the associated marker loci could not be assessed. Therefore, deviations in the negative direction from the check could be due to a deterioration of linkage integrity in some cases.

Nevertheless, as with any applied plant breeding program, availability of resources places limitations on the size of the program and certain compromises must be made. This situation affects the magnitude and success of breeding programs based on marker techniques just as it does in programs using traditional approaches.

Although comparisons of the above responses with those that might be obtained with traditional backcrossing and selfing would be informative, resources were not available to conduct a concurrent traditional study. If a traditional backcrossing program had been conducted without any field testing until the final testcrosses were made, it seems reasonable to expect that the number of testcrosses inferior to the check would be equal to (or perhaps, even greater than) the number of superior yielding testcrosses. Personal communications with corn breeders have indicated that attempts to improve B73 for traits such as yield using traditional methods have frequently met with failure. Also, with a traditional breeding approach, the time required would probably be at least double to achieve the level of homozygosity that was achieved with three backcrossing generations and two selfing generations using the DNA-based marker approach.

7. Marker-facilitated introgression (backcrossing)

Although the preceding section outlined an investigation that demonstrated the use of marker-facilitated backcrossing, the procedures used were probably more complex than will be encountered in many plant breeding strategies. Several of the important traits that must be manipulated by plant breeders are more simply inherited than grain yield but still may involve the expression of several genes. For example, disease resistance frequently is controlled by only a few genetic factors. However, for many diseases resistance is considered to be a multigenic ('semi-quantitative') trait. For example, Bubeck et al. (1993) have shown that resistance to gray leaf spot in maize is based on four or five genes.

Obviously, the first steps in any marker-based introgression program are the identification and mapping of the genes (more realistically, chromosomal segments) targeted for transfer to the desired line or strain. Experimental procedures for these steps have already been outlined in previous chapters. Once the appropriate analyses have been performed to identify the genes of interest in the resource (perhaps, exotic) strain, as well as linkages to resource-specific marker alleles, repeated backcrossing to the recipient line or cultivar – choosing in each cycle only those backcross progeny with desired linked marker alleles – will provide effective introgression of the desired genes of interest into the recipient line. As was demonstrated in the previous section, marker-assisted selection against unwanted chromosomal regions from the donor will expedite the introgression process.

Table 3 Frequency of a favorable allele after a given number of backcross generations, with and without selection for a linked marker allele or for a pair of markers bracketing the favorable allele, and with and without marker-assisted selection against the remaining donor genome (taken from Beckmann and Soller 1986)

Number of backcross generations	Marker-assisted selection				
	For favorable allele			Against remainder of donor genome	
	None	Single marker ^a	Marker bracket ^b	None	Full marker coverage ^c
	(Frequency of favorable allele)			(Proportion of recipient genome recovered)	
1	0.25	0.81	0.92	0.75	0.85
2	0.12	0.73	0.88	0.88	0.99
3	0.06	0.66	0.85	0.94	1.00
4	0.03	0.59	0.82	0.97	1.00
5	0.02	0.53	0.78	0.98	1.00
6	0.01	0.48	0.75	0.99	1.00

^a Proportion of recombination between marker allele and linked favorable allele = 0.10

^b Proportion of recombination between the two markers of the bracket = 0.40

^c Assuming two markers per chromosome

Table 3 (taken from Beckmann and Soller 1986) shows the frequency of a favorable allele following one to six backcross generations, with and without selection for a linked marker allele, including selection for a pair of bracketing marker alleles. It can be noted that the frequency of the introgressed favorable allele after three generations of backcrossing is 0.66 for the single marker and 0.85 for bracketing markers (with recombination of 0.40 between markers). These frequencies are in striking contrast to the 0.06 with no markers after three backcrosses and only 0.01 after six backcrosses. Also, with marker-assisted introgression, frequencies for the introgressed alleles are sufficiently high that two, three or even more alleles could be readily introduced and brought to fixation in a given breeding cycle. As stated by Beckmann and Soller (1986), 'Without marker assistance, a great many backcross products will have to be screened for the introduced trait, even in the case of one introduced allele, due to the extreme rarity of backcross products carrying desired exotic alleles.'

Using a novel marker-facilitated backcrossing scheme, Brown et al. (1989) transferred isozyme-marked segments from wild barley (*Hordeum spontaneum*) into an elite barley (*H. vulgare*) cultivar. Each of the 84 near-isogenic lines was then made homozygous for a single isozyme-marked segment with two selfing generations. After evaluating these lines in the field, they concluded that this novel backcrossing scheme was a useful approach for identifying QTLs for improving yield in divergent germplasm. A major advantage of this approach is that once a favorable QTL has been identified, it is already fixed in the line and the breeding work is essentially completed. In addition, lines with favorable QTL alleles can be easily maintained and then used for pyramiding several favorable QTL alleles into a single line.

8. Conclusions

The effectiveness of molecular-marker technology for identifying and mapping QTLs has been widely demonstrated in several crop plants. Also, the positive results from a limited number of marker-facilitated selection and introgression studies are encouraging for transferring desired genes between breeding lines and, thus, increasing the precision and efficiency of plant breeding. It is hoped that markers also should expedite the acquisition of important genes from exotic populations or from wild species.

In maize, for example, most mapping studies have been conducted in populations derived from domestic lines. Efforts in exotic maize populations have been less effective and frequently have met with considerable frustration (Koester 1992). This has been particularly true in the use of RFLPs because of the large number of marker variants and multiple banding patterns that have been difficult to interpret. Differentiating between multiple alleles at a single marker locus versus alleles associated with duplicate loci frequently is nearly impossible. Unfortunately, this situation presents a major limitation to the use of exotic germplasm as a vast source of new genes in maize breeding.

In most QTL identification studies, rather stringent probability levels have been set so that there is a low risk in making Type I errors (i.e., false positives). Thus, only QTLs with major effects are identified as being significant. It should be pointed out that these would be the QTLs with high heritabilities, are easily manipulated using traditional breeding practices, and may already be fixed in many breeding lines. It may prove to be more productive, therefore, to use marker technology as a means for placing greater emphasis on those QTLs (or chromosomal regions) that show only relatively minor effects.

There have been some attempts to reduce the size of the regions identified as containing major QTLs through 'fine-mapping' such as the study reported by Paterson et al. (1990). Many have envisioned this approach as an initial step in identifying single genes that might ultimately be manipulated using transformation (recombinant DNA) technology. One might speculate, however, that this approach could be counter-productive if these major QTLs contain a number of genes that have evolved as highly integrated complexes over many cycles of selection. For example, the region on the short arm of chromosome 5, bracketed by isozyme markers *Amp3* and *Pgm2*, has been found to have a very significant effect on grain yield in several maize populations (Stuber 1992; Stuber et al. 1992). This region is being targeted in some studies for fine-mapping. It will be of interest to determine whether chromosomal segments such as this can be further improved with marker technology. In practical breeding programs, however, manipulation of large segments (such as the one identified in maize on chromosome 5) should be much simpler than extracting and manipulating individual genes.

It should be stressed that little is known regarding the stability of QTL alleles when transferred to different genetic backgrounds and when evaluated in varying environments. Tanksley and Hewitt (1988) illustrated the potential dangers of establishing breeding programs based on associations of markers with quantitative traits prior to evaluations of the identified factors in appropriate genetic backgrounds.

Each new investigation using DNA-based marker technology as a tool for plant geneticists and plant breeders adds evidence to the projected role of markers not only for identifying useful genes (or chromosomal segments) in various germplasm sources but also for transferring these genes into desired cultivars or lines. As laboratory analyses become more automated, the cost (now one of the major deterrents in the use of marker technology) will decrease and the use of DNA-based markers for plant improvement will be greatly expanded in the future.

9. References

- Abler, B.S.B., Edwards, M.D. and Stuber, C.W. (1991) Isoenzymatic identification of quantitative trait loci in crosses of elite maize hybrids. *Crop Sci.* 31: 267–274.
- Beckmann, J.S. and Soller, M. (1986) Restriction fragment length polymorphisms in plant genetic

- improvement Oxford Surv Plant Mol Cell Biol 3 197-250
- Brown, A H D, Munday, J and Oram, R N (1989) Use of isozyme-marked segments from wild barley (*Hordeum spontaneum*) in barley breeding Plant Breed 100 280-288
- Bubeck, D M, Goodman, M M, Beavis, W D and Grant, D (1993) Quantitative trait loci controlling resistance to gray leaf spot in maize Crop Sci 32 838-847
- Comstock, R E (1978) Quantitative genetics in maize breeding In D Walden (ed), Maize Breeding and Genetics, pp 191-206 Wiley, New York
- Darrah, L L and Hallauer, A R (1972) Genetic effects estimated from generation means in four diallel sets of maize inbreds Crop Sci 12 615-621
- Dudley, J W, Saghai Maroof, M A and Rufener, G K (1992) Molecular marker information and selection of parents in corn breeding programs Crop Sci 32 301-304
- Duvick, D N (1984) Genetic contributions to yield gains of U S hybrid maize, 1930 to 1980 In W R Fehr (ed), Genetic Contributions to Yield Gains of Five Major Crop Plants, pp 15-47 Crop Science Society of America, Madison, WI
- East, E M and Hayes, H K (1911) Inheritance in maize Conn Agr Expt Sta Bull 167
- Edwards, M D, Stuber, C W and Wendel, J'F (1987) Molecular-marker-facilitated investigations of quantitative-trait loci in maize I Numbers, genomic distribution, and types of gene action Genetics 116 113-125
- Edwards, M D, Helentjaris, T, Wright, S and Stuber, C W (1992) Molecular-marker-facilitated investigations of quantitative trait loci in maize IV Analysis based on genome saturation with isozyme and restriction fragment length polymorphism markers Theor Appl Genet 83 765-774
- Emerson, R A and East, E M (1913) The inheritance of quantitative characters in maize Nebr Agr Expt Sta Res Bull 2
- Frei, O M, Stuber, C W and Goodman, M M (1986a) Use of allozymes as genetic markers for predicting performance in maize single cross hybrids Crop Sci 26 37-42
- Frei, O M, Stuber, C W and Goodman, M M (1986b) Yield manipulation from selection on allozyme genotypes in a composite of elite corn lines Crop Sci 26 917-921
- Gonella, J A and Peterson, P A (1978) Isozyme relatedness of inbred lines of maize and performance of their hybrids Maydica 23 55-61
- Guse, R A, Coors, J G, Drolsom, P N and Tracy, W F (1988) Isozyme marker loci associated with cold tolerance and maturity in maize Theor Appl Genet 76 398-404
- Hadjinov, M I, Scherbak, V S, Benko, N I, Gusev, V P, Sukhorzheuskaya, T B and Voronova, L P (1982) Interrelationships between isozyme diversity and combining ability in maize lines Maydica 27 135-149
- Hayes, H K (1952) Development of the heterosis concept In J Gowen (ed), Heterosis, pp 49-65 Iowa State College Press, Ames, IA
- Heidrich-Sobrinho, E and Cordeiro, A R (1975) Codominant isoenzymatic alleles as markers of genetic diversity correlated with heterosis in maize (*Zea mays* L) Theor Appl Genet 46 197-199
- Hunter, R B and Kannenberg, L W (1971) Isozyme characterization of corn (*Zea mays* L) inbreds and its relation to single cross hybrid performance Can J Genet Cytol 13 649-655
- Kahler, A L (1985) Associations between enzyme marker loci and agronomic traits in maize Proc 40th Annual Corn and Sorghum Industry Research Conf, American Seed Trade Assoc 40 66-89
- Koester, R P (1992) Identification of quantitative trait loci controlling maturity and plant height using molecular markers in maize (*Zea mays* L) Ph D thesis, North Carolina State University, Raleigh, NC
- Lande, R and Thompson, R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits Genetics 124 743-756
- Lamkey, K R, Hallauer, A R and Kahler, A L (1987) Allelic differences at enzyme loci and hybrid performance in maize J Hered 78 231-234
- Melchinger, A E, Lee, M, Lamkey, K R and Woodman, W L (1990) Genetic diversity for

- restriction fragment length polymorphisms Relation to estimated effects in maize inbreds *Crop Sci* 30 1033–1040
- Paterson, A H , Deverna, J W , Lanini, B and Tanksley, S D (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato *Genetics* 124 735–742
- Paterson, A H , Damon, S , Hewitt, J D , Zamir, D , Rabinowitch, H D , Lincoln S E , Lander, E S and Tanksley, S D (1991) Mendelian factors underlying quantitative traits in tomato Comparison across species, generations, and environments *Genetics* 127 181–197
- Peng, J Y , Glaszmann, J C and Virmani, S S (1988) Heterosis and sozyme divergence in indica rice *Crop Sci* 28 561–563
- Pollak, L M , Gardner, C O and Parkhurst, A M (1984) Relationships between enzyme marker loci and morphological traits in two mass selected maize populations *Crop Sci* 24 1174–1179
- Russell, W A (1984) Agronomic performance of maize cultivars representing different eras of maize breeding *Maydica* 29 375–390
- Smith, J S C and Smith, O S (1989) The use of morphological, biochemical, and genetic characteristics to measure distance and to test for minimum distance between inbred lines of maize (*Zea mays* L.) (Mimeo of paper presented at UPOV Workshop, Versailles, France, October 1989) Pioneer Hi-Bred International, Inc , Johnston, IA
- Stuber, C W and Edwards, M D (1986) Genotypic selection for improvement of quantitative traits in corn using molecular marker loci *Proc 41st Annual Corn and Sorghum Industry Research Conf , American Seed Trade Assoc* 41 70–83
- Stuber, C W and Moll, R H (1972) Frequency changes of isozyme alleles in a selection experiment for grain yield in maize (*Zea mays* L.) *Crop Sci* 12 337–340
- Stuber, C W and Sisco, P H (1991) Marker-facilitated transfer of QTL alleles between elite inbred lines and responses in hybrids *Proc 46th Annual Corn and Sorghum Industry Research Conf , American Seed Trade Assoc* 46 104–113
- Stuber, C W , Moll, R H , Goodman, M M , Schaffer, H E and Weir, B S (1980) Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays* L.) *Genetics* 95 225–236
- Stuber, C W , Goodman, M M and Moll, R H (1982) Improvement of yield and ear number resulting from selection at allozyme loci in a maize population *Crop Sci* 22 737–740
- Stuber, C W , Edwards, M D and Wendel, J F (1987) Molecular marker-facilitated investigations of quantitative trait loci in maize II Factors influencing yield and its component traits *Crop Sci* 27 639–648
- Stuber, C W , Lincoln, S E , Wolff, D W , Helentjaris, T and Lander, E S (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers *Genetics* 132 823–839
- Tanksley, S D and Hewitt, J (1988) Use of molecular markers in breeding for soluble solids content in tomato – a re-examination *Theor Appl Genet* 75 811–823

6. Nuclear DNA markers in systematics and evolution

RICHARD WHITKUS¹, JOHN DOEBLEY² and JONATHAN F. WENDEL³

¹ Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, U.S.A.;

² Department of Plant Biology, University of Minnesota, St. Paul, MN 55108, U.S.A.; ³ Department of Botany, Iowa State University, Ames, IA 50011, U.S.A.

Contents

1. Introduction	116	3. Biological questions	124
2. Technical aspects and considerations	117	3.1. Evolutionary pattern	124
2.1. Restriction fragment length polymorphisms	117	3.2. Evolutionary process	130
2.2. Randomly amplified polymorphic DNA	119	3.2.1. Genome evolution	130
2.3. Dispersed repetitive DNA	120	3.2.2. Gene flow and introgression	133
2.4. Data collection and analysis	121	3.2.3. Evolution of morphological traits	134
2.4.1. Population sample sizes	121	4. Conclusion	135
2.4.2. Number of markers to assay	122	5. References	136
2.4.3. Interpreting gels and autoradiographs	123		

Abbreviations: bp – base pair(s); cpDNA – chloroplast DNA; drDNA – dispersed repetitive DNA; kb – kilobase(s); mtDNA – mitochondrial DNA; PCR – polymerase chain reaction; QTL – quantitative trait locus; RAPD – random amplified polymorphic DNA; RFLP – restriction fragment length polymorphism

1. Introduction

Systematic biology studies the organization and origin of biological diversity. Systematists analyze patterns of variation among the products of evolution (species) and attempt to infer the processes involved in evolutionary change from the observed pattern. A variety of approaches have traditionally been used to infer systematic or evolutionary relationships among plant species including morphology, anatomy, cytology, breeding systems, cross-compatibility, secondary compounds (e.g. flavonoids), and isozymes. Each of these approaches has its limitations. For example, although isozymes are useful in

analyzing variation within and among closely related species, they are generally not applicable to the study of more distantly related taxa (Crawford 1990). In contrast, morphological comparisons can be made among widely divergent species, but our ignorance of the underlying mechanism of morphological evolution creates uncertainty about the extent to which morphological similarity accurately reflects evolutionary relationships

The limitations and uncertainties surrounding traditional sources of systematic evidence have created a continuing search for new and more robust sources of data. In this regard, recent advances in molecular biology have provided a variety of new tools applicable in systematic research. Most notably, restriction site analysis of chloroplast DNA (Palmer et al. 1988) and nuclear ribosomal genes (Hamby and Zimmer 1992) have been successfully applied in many studies over the past decade. Direct analysis of nucleic acid sequences has also increased in systematics (e.g. Miyamoto and Cracraft 1991; Kim et al. 1992). In this chapter, we examine another source of molecular data applicable to systematics – nuclear DNA markers. Nuclear DNA markers include restriction fragment length polymorphisms (RFLP), random amplified polymorphic DNA (RAPD), and dispersed repetitive DNA (drDNA). As will be discussed, these markers provide a potentially unlimited number of loci distributed over the entire nuclear genome and furnish a powerful means for detecting variation within and among species. We begin with a review of some technological considerations surrounding nuclear DNA markers and then discuss their application to systematic research.

2. Technical aspects and considerations

2.1. Restriction fragment length polymorphisms

The digestion of genomic DNA with a restriction endonuclease produces a population of DNA (restriction) fragments of varying lengths. Standard molecular techniques including agarose gel electrophoresis, Southern transfers and filter hybridizations with radioactively labeled probes (cloned DNA fragments) enable one to visualize the individual restriction fragments homologous to the probe (Sambrook et al. 1989). When the genomes of two or more individuals are compared by these methods, it is possible to detect differences or polymorphisms in the lengths of their restriction fragments (RFLPs). Such polymorphisms can occur as a result of base-pair changes in the recognition sites for the restriction enzyme or rearrangements (e.g. insertion/deletions) in the region of the genome surrounding the cloned DNA fragment used as the probe.

Cloned probes for RFLP analysis may be obtained from either cDNA or genomic DNA libraries (Evola et al. 1986; Helentjaris et al. 1986). From these libraries, one selects probes that represent the unique sequence and/or low-copy-number portions of the nuclear genome (Landry et al. 1987; Landry and

Michelmore 1987). A unique sequence probe will hybridize to a single restriction fragment (single locus), provided there are no recognition sites for the restriction enzyme within the portion of the genome homologous to the probe. Low-copy-number probes may hybridize to as many as five loci. As locus number increases, the ability to resolve and accurately score the fragments diminishes. Thus, researchers engaged in RFLP analyses normally select probes that hybridize to only one or two loci.

Several technique-related factors may influence the amount of RFLP diversity that one detects within and among species. First, it has been shown that various restriction enzymes have different efficiencies for revealing RFLPs (Landry et al. 1987; McCouch et al. 1988; Miller and Tanksley 1990a). The enzymes *DraI*, *EcoRI*, *EcoRV*, *HindIII*, and *XbaI* often reveal high levels of polymorphism. Second, cDNA probes appear to reveal greater polymorphism than random genomic probes, at least in some species (Landry et al. 1987; Miller and Tanksley 1990a).

Available evidence from surveys of RFLPs within species indicates that RFLPs normally involve changes in fragment size and not the loss of fragments homologous to the probe DNA (Evola et al. 1986; Helentjaris et al. 1985; McCouch et al. 1988). For this reason, RFLPs are generally codominantly inherited with 1:2:1 segregation ratios in F_2 populations (Beckmann and Soller 1983; Helentjaris et al. 1986).

RFLP analysis has several important advantages. First, RFLPs are commonly and easily detected both within and among species. Second, the use of filter hybridization procedures provides assurance that fragments detected in different individuals or species contain homologous sequences, although, for duplicated loci, one may often have difficulty distinguishing orthologous from paralogous fragments. Third, RFLP linkage maps constructed with one mapping population of a species will be useful not only for other populations of the same species but also for other closely related species (Bonierbale et al. 1988; Paterson et al. 1991). Fourth, the codominant nature of RFLP loci makes them better suited to mapping quantitative trait loci (QTLs) than dominant markers such as RAPDs (see below). Fifth, the ability to score RFLPs as alleles at a locus enables one to take advantage of a variety of statistical methods developed for estimating heterozygosity, gene flow, population differentiation, and genetic distances (Gottlieb 1981; Nei 1987; Hartl and Clark 1989; Weir 1990a,b).

Unfortunately, these advantages do not come without cost since the analysis of RFLP variation among plants has a formidable set of prerequisites. Reasonably large quantities of genomic DNA (100 micrograms or more) must be isolated from each individual plant, a library of cloned DNA sequences from the species under study or a related species must be available, and researchers must be familiar with the various laboratory procedures outlined above. These requirements make RFLP analysis more labor intensive and costly than isozyme analysis (Beckmann and Soller 1983).

2.2. *Randomly amplified polymorphic DNA*

Randomly amplified polymorphic DNA (RAPD) markers represent amplification products from a polymerase chain reaction (PCR) utilizing arbitrary primers and genomic DNA (Williams et al. 1990). The primers are normally 10 bp in length and have a GC content of at least 50%. Unlike typical PCR reactions, a single, rather than a pair of primers, is used for amplification. Two correctly-oriented primer sites must exist within a limited distance (typically under 3000 bp) in the genomic DNA for amplification to be successful. Amplification will not occur if primer sites are much further apart in the genome. Amplification products are visualized by separation on agarose gels and staining with ethidium bromide.

Most variation among individuals for RAPDs probably arises from base pair substitutions or insertion/deletions that modify (or eliminate) the primer site, or insertions in the genomic sequence that separate the primer sites to a distance that will not permit amplification (Williams et al. 1990). Each of these changes results in presence/absence of a particular RAPD fragment. For this reason, RAPDs are normally inherited in a dominant fashion in F_2 populations, exhibiting 3:1 segregation ratios. In rare cases, variation at a RAPD locus may result in the alteration of the DNA fragment size rather than fragment present/absent; codominant inheritance may be expected in these instances. If codominant expression of RAPDs is desired, it may also be obtained when an invariant RAPD fragment is polymorphic for an internal restriction site (Williams et al. 1990; Halward et al. 1991). Digestion of the fragment with the appropriate restriction enzyme followed by gel electrophoresis will reveal the polymorphism.

Amplification of genomic DNA with a single primer often produces multiple RAPD fragments. This results whenever there are multiple sites in the genome that allow amplification. However, without formal genetic analyses, it is not possible to determine whether a two banded RAPD phenotype represents two loci or two distinct alleles at a single locus. Also, the commonly used 10 bp primers may amplify unrelated regions of the genome of two related species or even two individuals of a single species. Thus, homologies among fragments are unknown even when the fragments are produced by amplification with the same primer.

There are several advantages of RAPDs relative to RFLPs and isozymes. First, RAPDs do not require labor intensive and costly procedures such as Southern transfers, filter hybridizations, and autoradiography. Second, it is not necessary to construct or maintain a genomic library. Third, RAPDs require far smaller quantities of genomic DNA (a few hundred nanograms) than needed for RFLP analysis. Fourth, RAPDs, like RFLPs but unlike isozymes, provide an essentially unlimited number of markers throughout the genome. Fifth, preliminary data suggest that RAPDs reveal high levels of polymorphism even within and among species that show little RFLP or isozyme polymorphism (Halward et al. 1991; Paran et al. 1991; Van Heusden and Bachmann 1992).

There are several concerns surrounding the use of RAPDs in systematic research. First, Mendelian dominance of the majority of the loci eliminates the ability to distinguish heterozygotes which is a desirable capability for many methods of data analysis. Second, primers may amplify different regions of the genome in different individuals or species. Consequently, homologies are uncertain, leading to several problems including the fact that linkage maps developed in one mapping population may not be useful in another, and the inability to accurately use standard statistics for measuring heterozygosity, gene flow, population differentiation, and genetic distances. Third, while shared presence of a RAPD fragment is likely to represent a true homology (at least between closely related entities), shared absence of a RAPD fragment cannot be assumed homologous among individuals since loss of the fragment can occur by numerous independent events. Finally, PCR can be sensitive to factors such as primer and substrate concentrations and impurities in DNA preparations, indicating that rigorously controlled experimental conditions are necessary for reliable and reproducible results.

2.3. Dispersed repetitive DNA

Eucaryotic genomes typically possess short sequence motifs, two to 15 base pairs in length, repeated many times in tandem clusters that are dispersed throughout the genome (Tautz and Renz 1984; Flavell 1986; Lewin 1990). These sequences, which we refer to as dispersed repetitive DNA (drDNA), represent a major component of heterochromatin and satellite DNA (Lewin 1990). Detection of variation in drDNA can be accomplished by restriction digests of genomic DNA, Southern transfers, filter hybridization and autoradiography. Cloned DNA for hybridization probes are available from several sources. First, short repeat sequences have been isolated from the human genome (e.g. Jeffreys et al. 1985a,b) that hybridize to drDNA in many species including plants (Dallas 1988; Rogstad et al. 1988). Second, a portion of the M13 phage cloning vector similarly detects drDNA in a variety of species (Vassart et al. 1987; Ryskov et al. 1988; Zimmerman et al. 1989). Third, synthetic oligonucleotides containing a short sequence motif (e.g. GATA) repeated three or four times readily detect drDNA (Tautz and Renz 1984; Epplen 1988; Tautz 1989).

Variability at drDNA loci can arise from unequal exchange during recombination within long stretches of repetitive sequences or slippage during DNA replication, producing variation in the number of repeats at a locus (Smith 1976; Jones and Kafatos 1982; Walsh 1987). Length mutations at drDNA loci have been estimated to be as high as 0.4% per locus per gamete, resulting in per locus estimates of heterozygosity above 90% (Jeffreys et al. 1985a, 1988; Georges et al. 1990). Restriction enzyme digests and Southern hybridization with drDNA probes reveal a complex pattern of bands on the autoradiograph. The high molecular weight fragments are usually well-separated from one another but the low molecular weight fragments can form a large smear of poorly resolved bands. Because of the large number of bands

observed and the high degree of variability among individuals, each individual may possess a unique banding pattern or 'DNA-fingerprint' (Jeffreys et al. 1985b). Because drDNA loci are highly polymorphic for the number of repeat units, they have been called hypervariable regions (HVR) or variable number tandem repeat loci (VNTR) (Jeffreys et al. 1985a, 1988; Georges et al. 1990). The high level of polymorphism revealed by drDNA is its principal advantage for evolutionary studies.

Analysis of drDNA has some troublesome limitations. Because drDNA analysis reveals many restriction fragments from numerous loci, there is no accurate means of determining homology among bands on autoradiographs. This, in turn, severely restricts one's ability to treat similarity/dissimilarity in band phenotypes to rigorous statistical analyses (Lynch 1988). Furthermore, drDNA loci detected by some probes in some species may be confined to a few linked regions of the genome in which case they may not provide an accurate measure of overall genetic similarity (Jeffreys et al. 1987; Georges et al. 1990).

2.4. Data collection and analysis

It is beyond the scope of this review to provide a detailed discussion of sampling, experimental design and data analysis. For in depth discussion of these topics in phylogenetic and systematic research, we refer readers to several excellent sources (Nei 1987; Felsenstein 1988; Baverstock and Moritz 1990; Weir 1990a,b). However, we feel that some issues relevant to the use of nuclear DNA markers in studies of genetic diversity and phylogeny deserve comment.

2.4.1. Population sample sizes

In studies of phylogeny and genetic diversity, there are numerous factors that determine how many individuals per population or populations per species one should analyze. These factors include the amount of ecogeographic diversity within the species under study, the size of natural populations of the species, the level of polymorphism that exists within and among populations, and the specific research question being addressed. For example, extensive sampling within species with low levels of diversity is a less than optimal allocation of resources. Thus, for cpDNA, which generally shows little variation within species and populations, most systematists have analyzed relatively few (1 to 3) populations per species and generally only a single individual per population (but see Soltis et al. 1992a). This is a reasonable strategy when the goal is to assess phylogenetic relationships among species. Similarly, extensive sampling for allozyme variation within populations of inbreeding species, which generally possess low levels of diversity, may not improve one's estimates of intrapopulational diversity (Gottlieb 1981). Conversely, populations of outbreeding species tend to possess considerable allozyme variation, requiring larger populational samples for accurate estimation of allele frequencies (Crawford 1983).

Although surveys for nuclear DNA marker variation in plant populations

are relatively few to date, available evidence indicates that such markers are often highly polymorphic within species (see below). For this reason, estimates based on one or a few population samples may not accurately reflect diversity within a species. Similarly, levels of polymorphism for nuclear DNA markers within populations of outbreeding species are likely to be high so that sampling a single individual per population, as is often done for cpDNA, is not likely to accurately represent variation within the population.

2.4.2. *Number of markers to assay*

In studies of isozyme variation in plants, the number of loci assayed is often determined by a technical limitation, namely the maximum number of loci that can be electrophoretically resolved. This is rarely more than 30. This limitation clearly does not apply to RFLP and RAPD markers for which the potential number of markers is essentially unlimited. Accordingly, a relevant question becomes how many molecular marker loci should one assay? To our knowledge, there is little empirical or theoretical evidence to provide a reliable answer for this question. Nei (1978) has suggested that ideally more than 50 loci should be used to accurately estimate heterozygosity. Nei et al. (1983) suggested that more than 30 loci should be used when estimating phylogenies from gene frequency data. The experience of one of us (J.W.) suggests that estimates of population genetic parameters, such as heterozygosity, may be stabilized when the number of RFLP loci analyzed reaches 50 (although clearly this a preliminary estimate). One advantage of employing larger numbers of loci is that accurate estimates of genetic parameters can be obtained even if few individuals are assayed (Nei 1978; Weir 1990a). In practice, individual researchers might consider analyzing their data at set intervals to determine when the addition of more markers ceases to provide significant new information.

When surveying RFLP diversity among and within species, a relevant concern is the appropriate number of restriction enzymes that one should employ with each probe. One might independently digest genomic DNAs with three different enzymes and then probe each of the resulting blots with 10 probes (30 total hybridizations). Alternatively, one might use a single enzyme and 30 probes. The amount of laboratory work is essentially equivalent, but the amount of useful information recovered may not be. For example, if correlations occur among the banding patterns revealed by a single probe with multiple enzymes, then it would be a better investment of resources to use more probes, each with only a single enzyme. A correlation among the patterns observed with different enzymes is likely to occur when the molecular events underlying the differences are insertion/deletions rather than restriction site changes. Results from one of our labs (J.D.) with sorghum indicate that the patterns of variation revealed by different enzymes with a single probe are often correlated. While it is not known if these correlations will prove to be typical, the best strategy may be to use many probes each with a single enzyme in cases where the number of available probes is not limiting.

2.4.3. *Interpreting gels and autoradiographs*

As mentioned above, there are different means of encoding the banding patterns one sees on gels and autoradiographs for statistical analysis. First, with RFLPs, one may know that all bands visualized with a single probe represent different states at a single location in the genome. In this case, it is possible to treat each band or set of bands as an allele at a locus. The application of this approach is easiest when the probes represent unique sequences (single loci) and the species being compared are not differentiated by events that produce differences in gene copy numbers such as polyploidy, aneuploidy or segmental duplication. Interpretation of autoradiographs in a locus/allele model is also simpler for predominantly self-pollinating species as the likelihood of confusing heterozygotes with two-banded alleles is reduced. One advantage of using the locus/allele model is the ability to analyze the resulting data with a variety of available statistical techniques (Nei 1987; Gottlieb 1981; Hartl and Clark 1989; Weir 1990a,b).

An alternate approach to encoding banding patterns would be to simply score presence versus absence of each band. Although this approach is straightforward, the data are less amenable to analysis with standard measures of genetic diversity and distance. Thus, estimates of relatedness must be inferred from the proportion of shared bands (Lynch 1988, 1990, 1991). The shared presence of a homologous RAPD, RFLP or drDNA fragment can also be used as character states in parsimony phylogenetic analyses (Swofford and Olsen 1990).

Above we discussed the degree of certainty one can have that two fragments visualized on a gel or autoradiograph represent homologous sequences. Certainty of sequence homology is greatest for unique sequence probes used to detect RFLPs and much less for RAPDs and drDNA. The question of homology also exists at another level, i.e. are DNA fragments, whether RFLPs, RAPDs or drDNAs, of the same size (electrophoretic mobility) fully homologous as opposed to just containing some sequence homology. For example, a 5 kb RFLP fragment in species A may mutate to a 6 kb fragment by the loss of a restriction site in species B and by a 1 kb insertion in species C. Although species B and C now share a 6 kb fragment that hybridizes to the same unique sequence RFLP probe, these fragments are not truly homologous.

The probability of convergence to similar RFLP fragment sizes is apt to be more critical at some levels of comparison than others. For example, among closely related samples (individuals within a species), there has been relatively little evolutionary time for convergent evolution to have occurred. However, when comparisons of nuclear DNA markers are made among species and genera, there is an increased probability that fragments of similar electrophoretic mobility are not fully homologous.

3. Biological questions

Plant systematics has historically benefited from the advent and application of biochemical and molecular sources of evidence (reviewed in Crawford 1990). This process is likely to continue as the relatively newer methods discussed in this review become broadly applied to a variety of systematic questions. Manuscripts utilizing RFLPs, RAPDs and drDNAs are appearing at an increasing pace, although the volume of published literature is as yet relatively modest. Nonetheless, it is already apparent that these tools will significantly impact plant systematics in at least two, not wholly separable ways. First, the ready availability of a large number of genetic markers promises to increase resolution for a variety of applications that represent the traditional concerns of systematists, i.e. understanding the pattern of organismal diversity and its hierarchical organization. Second, the ability to generate densely marked linkage maps of entire genomes, which was heretofore not possible for most taxa, represents a powerful approach that promises to yield novel insights into a number of evolutionary phenomena and mechanisms, such as speciation processes, genome evolution and the genetic basis of morphological transformations. Example applications of RFLPs, RAPDs and drDNAs to these two broad foci of plant systematic biology are presented below under the headings 'Evolutionary Pattern' and 'Evolutionary Process', although this division is to a certain extent artificial. For example, specific evolutionary transformations in morphology or genome structure are justifiably postulated only when phylogenetic relationships among the relevant taxa are well-understood. Similarly, studies of gene flow (process) contribute to an understanding of patterns of genetic and organismal diversity.

3.1. *Evolutionary pattern*

Systematics is fundamentally concerned with understanding diversity and its hierarchical organization. Consequently, the principal activity of most systematists has been and continues to be classification. Traditionally, classification has been based primarily on morphological characteristics, although in many cases these data have been supplemented by evidence from a variety of other sources such as cytogenetics and chemotaxonomy. Although often readily obtained, the number of morphological characters available for analysis is typically limited to a relatively small number, at least in comparison with molecular characters. In addition, the genetic basis of most morphological variation is generally unknown, although qualitative differences in morphology can be under relatively simple genetic control (Bachmann 1983; Hilu 1983; Gottlieb 1984). Because molecular characters provide a rich source of genetically interpretable data, often entailing hundreds or even thousands of different characters, molecular approaches to systematic problems have enjoyed increasing popularity during the last decade.

For purposes of classification and phylogeny, the most frequently and

fruitfully applied tools have been restriction site analysis of chloroplast DNA and DNA sequencing of individual genes (recently reviewed by Crawford 1990; Hillis and Moritz 1990; Soltis et al. 1992b). Notwithstanding the insights gained from such studies, cpDNA restriction site analysis and gene sequencing suffer from several potential problems, which collectively or individually may yield 'gene trees' that depart in topological detail from the (true but unknown) 'organismal trees'. In studies that involve single gene variation from several or more taxa, one must consider the possibilities of lineage sorting (differential lineage survival of ancestral polymorphisms; Nei 1987; Ball et al. 1990), and inclusion of a mixture of paralogous and orthologous sequences from the study taxa (Doyle 1992). Phylogenetic studies based on cpDNA restriction site analysis are subject to several potential difficulties that arise from the use of uniparentally inherited, non-recombinant molecules, i.e. lineage sorting, inter-taxon cytoplasmic introgression (reviewed in Rieseberg and Brunsfeld 1992; Rieseberg and Soltis 1991; Soltis et al. 1992a), and non-independence of characters (Doyle 1992).

Although the seriousness of the above-mentioned phenomena is arguable with respect to any specific phylogenetic study, they are expected to be less of a concern in investigations that employ numerous, nuclear-encoded molecular markers such as RFLPs or RAPDs, where each marker-locus may be considered a more or less independent estimator of phylogeny. Because of this, lineage sorting and cytoplasmic introgression are relatively minor concerns. This is not to suggest, however, that the RFLP or RAPD approach is the method of choice for phylogenetic inference; in fact, problems of homology determination across taxa may be insurmountable in some instances, especially with RAPDs and drDNAs; when this is the case, phylogeny reconstruction is inadvisable. Perhaps for this reason, molecular markers have found only limited application to phylogeny reconstruction and classification above the rank of species. The few published studies, all using RFLP analysis, are on crop species and their wild relatives, the most extensive involving tomatoes (Miller and Tanksley 1990b), cole crops (Song et al. 1988, 1990, 1991) and lettuce (Kesseli et al. 1991).

Miller and Tanksley (1990b) examined phylogenetic relationships among eight species of *Lycopersicon* using 40 RFLP clones covering the majority of the tomato genome. A primary result is that their dendrogram contained two clusters, corresponding to clades of self-incompatible (*L. peruvianum*, *L. pennelii* and *L. hirsutum*) and self-compatible (*L. esculentum*, *L. pimpinellifolium*, *L. cheesmanii*, *L. parviflorum*, and *L. chmielewskii*) species. The red-fruited species (*L. esculentum*, *L. pimpinellifolium*, *L. cheesmanii*) formed a separate cluster within the self-compatible species. This depiction of relationships, which closely matches those derived from morphology and crossing studies, differs from two previous molecular phylogenies based on organellar DNA (cpDNA – Palmer and Zamir 1982; mtDNA – McClean and Hanson 1986). These latter studies did not resolve species into self-compatible and self-incompatible clades, nor did they uniformly support a clade of red-

fruited species. While it remains unclear what the 'correct' phylogenetic topology is for *Lycopersicon*, this example is instructive in that the phylogenies generated from organellar DNA and nuclear molecular markers are different.

Phylogenetic analysis of *Brassica* RFLPs (Song et al. 1988, 1990, 1991) has led to several noteworthy results. First, genomes within diploid *Brassica* species appear to have evolved by increasing aneuploidy (see Evolutionary process, below). Second, *B. rapa* is suggested to have been domesticated at least twice, with turnip originating from European wild types, whereas pak choi, Chinese cabbage and *narinosa* are postulated to have been derived from wild East Asian populations. Third, all cultivated forms of *B. oleracea* (including broccoli, cabbage, thousand-head kale, Chinese kale, kohlrabi and cauliflower) appear to be monophyletic, providing evidence for a single domestication event in the history of this important crop.

Insights into the relationship between lettuce (*Lactuca sativa*) and prickly lettuce (*L. serriola*) were obtained by Kesseli et al. (1991) using 55 RFLP clones. They showed that *L. virosa*, *L. indica*, *L. saligna*, and *L. perennis* clustered separately from *L. serriola* and *L. sativa*. Within the *L. serriola*-*L. sativa* cluster, some accessions of *L. serriola* formed a sister cluster to the *L. sativa* complex, while others were scattered within the complex. The pattern of clustering found with the accessions, and a dichotomous distribution of high frequency 'alleles' in cultivars of lettuce, led to the suggestion that cultivated lettuce may have had a polyphyletic origin.

In addition to these few examples of classification and phylogeny reconstruction at the species level and above, RFLPs have been used in efforts to reveal the parentage of polyploids in a number of taxa (e.g. Song et al. 1988; Hosaka et al. 1990; Kochert et al. 1991; Jena and Kochert 1991). Each of these studies was relatively limited in scope, both in terms of thoroughness of taxon sampling and with respect to the number of probes utilized.

Much more common has been the use of molecular markers in describing patterns of infraspecific diversity. At least two factors appear to be responsible for the larger volume of molecular marker literature at the sub-specific than supra-specific rank. The first is that molecular markers are more likely to be homologous in a sampling of closely related taxa, populations or accessions than among more distantly related entities. While this principle applies to all characters, molecular or otherwise, some techniques appear to be more sensitive to the relationship between relatedness and confidence in homology assessment than others. As noted above (see Technical Considerations), confidence that similar RAPD and drDNA fragments are homologous is expected to be very low above the level of species, while this problem is not as acute for RFLPs. The second reason for the more frequent application of molecular markers to infraspecific problems is sociological/historical. For the most part, these techniques filtered into the systematics community through agronomic scientists, who were principally interested in the development of detailed genetic maps for various crops and in obtaining a more powerful set of markers for estimating levels and patterns of genetic diversity. This historical bias towards

crop plants, and to a lesser extent their wild relatives, is evidenced in the published literature, which consists, with rare exceptions, of examples from agricultural science.

One of the more common applications at the infraspecific level has been the description of patterns of relationships among populations or accessions based on RFLPs. Typically, several to many populations or accessions are examined for RFLP 'alleles' and the data are subjected to phenetic and/or cladistic analysis. Examples include *Arachis hypogaea* (Kochert et al. 1991), *Brassica* spp. (McGrath and Quiros 1992; Song et al. 1988, 1990, 1991), *Cucumis melo* (Neuhausen 1992), *Gossypium hirsutum* (C. Brubaker and J. Wendel, in press), *Lycopersicon* spp. (Miller and Tanksley 1990b), *Medicago sativa* (Brummer et al. 1991), *Oryza* spp. (Jena and Kochert 1991), *Phaseolus vulgaris* (Gepts 1993), *Sorghum bicolor* (Aldrich and Doebley 1992), and *Vicia faba* (Van de Ven et al. 1990). The increased resolution obtained with RFLPs, as compared to isozyme analysis, is exemplified by the study of 56 accessions of wild and cultivated *Sorghum bicolor* (Aldrich and Doebley 1992). Not only was greater diversity detected in the RFLP data set, but the variation present was distributed in a systematically and geographically more meaningful fashion. Specifically, the RFLP data lend powerful support to the suggestion that cultivated sorghum was domesticated from wild sorghum of central-northeastern Africa.

A second application at the infraspecific level has been in discriminating individuals via molecular marker 'fingerprinting'. RFLPs, RAPDs and drDNAs are particularly well suited to problems requiring genotyping of individuals, with objectives ranging from varietal identification or cultivar discrimination to the analysis of the multilocus genotypic structure of wild plant populations. Example applications for purposes of varietal identification or discrimination include Smith and Smith (1991) on hybrid maize (RFLP), Graner et al. (1990) on barley (RFLP), Nybom (1990) on apples and cherries (drDNA), Wilde et al. (1992) on *Theobroma cacao* (RAPD), and Nybom et al. (1989) on blackberries and raspberries (drDNA). That drDNA fingerprints can provide high resolution genetic data for wild populations was demonstrated by Rogstad et al. (1991a) for clones of *Populus tremuloides* in Colorado. Using the M13 probe, trees from individual sites could be assigned to genets with high confidence, even from interdigitating, morphologically cryptic clones. In addition, drDNA variation was sufficiently great that the probability that two randomly selected individuals would have the same fingerprint was estimated to be approximately 1 in 3000. A similar magnitude of discrimination (1/1700) was reported for individuals of *Asimina triloba* (pawpaw) from a wide geographic sampling of sites (Rogstad et al. 1991b). Other examples from wild populations include Nybom and Rogstad (1990) on *Acer negundo* (drDNA), and Nybom and Schaal (1990) on *Rubus* (drDNA).

One outgrowth of the application of molecular markers to problems at the infraspecific level is that considerable information is being generated regarding levels of genetic variability in both wild populations and domesticated plants.

Most of these data are derived as incidental by-products of research not explicitly designed to assess genetic diversity. Nonetheless, data on genetic variability within and among populations may provide insights into the events and population characteristics that shape allelic distribution and evolutionary change. To date, most information on genetic diversity comes from allozyme analysis, which has shown important relationships between life history characteristics and historical factors with levels and patterns of genetic diversity (reviewed in Loveless and Hamrick 1984; Soltis and Soltis 1989; Crawford 1990; Hamrick and Godt 1990). Because the number of isozyme loci surveyed in plants has typically ranged from 10 to 30, the larger number of loci available with molecular markers should add resolution and increase confidence in estimates of genetic variability. Among the marker classes considered in this review, the most appropriate for genetic diversity estimation are RFLP surveys, because the data generated are readily transformed into locus/allele models.

As one might expect, every RFLP polymorphism survey to date has demonstrated that taxonomic rank is positively correlated with amount of genetic diversity. For example, cultivars or accessions within a species exhibit less polymorphism than different species or subspecies (Figdore et al. 1988; Miller and Tanksley 1990b; Van de Ven et al. 1990; Guo et al. 1991; Kesseli et al. 1991; Neuhausen 1992). This is illustrated by the data from Kesseli et al. (1991), who examined polymorphisms in 5 species of *Lactuca*. Each RFLP probe by restriction enzyme combination was treated as a locus and genetic distances (Nei 1987) were calculated between species. Each species was represented by at least 1 accession and 20 plants were sampled per accession. Mean genetic distances within species ranged from 0.19 to 0.68; genetic distances between species ranged from 0.87 to 2.71.

It is also apparent that the breeding system has a strong effect on levels of RFLP diversity. Self-pollinating plants such as tomato (Helentjaris et al. 1985; Miller and Tanksley 1990b), wheat (Liu et al. 1990), melon (Shattuck-Eidens et al. 1990), beans (Chase et al. 1991), barley (Heun et al. 1991), and peanut (Halward et al. 1991), generally have relatively limited RFLP diversity, although this is not always the case, e.g. *Arabidopsis* (Chang et al. 1988), rice (McCouch et al. 1988), lettuce (Landry et al. 1987; Kesseli et al. 1991), sorghum (Aldrich and Doebley 1992), and cotton (C. Brubaker and J. Wendel, in press). The highest levels of RFLP variability are found in outcrossing plants, for example, in maize (Helentjaris et al. 1985). Perhaps the most instructive comparison is provided by the data in Miller and Tanksley (1990b), who compared levels of RFLP variability (40 clones) between self-compatible and self-incompatible species in *Lycopersicon*. Self-compatible species had fewer restriction fragments, fewer novel restriction fragments and lower average infraspecific genetic distances than self-incompatible species. Interspecific genetic distances were also lower in the self-compatible than self-incompatible species groups. In total, the self-incompatible species were shown to possess nearly three times as much genetic variation than the self-compatible species (Miller and Tanksley 1990b).

These relationships between taxonomic rank and breeding system with RFLP diversity parallel the results of numerous isozyme surveys (Soltis and Soltis 1989; Crawford 1990; Hamrick and Godt 1990). What is not clear is whether RFLP loci are inherently more variable than isozyme loci in all plant species. The most thorough comparisons between RFLP and isozyme loci are in *Sorghum bicolor* (Morden et al. 1989, 1990; Aldrich and Doebley 1992; Aldrich et al. 1992) and *Gossypium hirsutum* (C. Brubaker and J. Wendel, in press). These two studies are also readily compared because diversity data for both isozymes and RFLPs were derived from similar sets of accessions (84 in *G. hirsutum*, 49 in *S. bicolor*), and because similar analytical methods were employed. In a sampling of 53 RFLP loci and 30 allozyme loci in both wild and cultivated *S. bicolor*, the RFLP loci were, on average, more variable, as indicated by all standard measures of genetic variability in both wild and cultivated accessions, i.e. percent polymorphic loci, P (85% vs. 70% for wild, 72% vs. 43% for cultivated), mean number of alleles per locus, A (3.34 vs. 2.13 for wild, 2.26 vs. 1.53 for cultivated) and mean panmictic heterozygosity, H (0.40 vs. 0.15 for wild, 0.28 vs. 0.09 for cultivated). In contrast, precisely the opposite pattern is evident in estimates from 171 RFLP loci and 50 allozyme loci in *G. hirsutum*, where RFLP loci are much less variable than allozyme loci (P = 26% vs. 46%; A = 1.27 vs. 1.70; H = 0.05 vs. 0.12). At present, the molecular phenomena underlying these differences between *Sorghum* and *Gossypium* are obscure.

A point that is somewhat tangential to this comparison concerns the number of RFLP loci that are necessary to reach stable estimates of genetic diversity. The only empirical data that even peripherally addresses this issue is for *Gossypium hirsutum* (C. Brubaker and J. Wendel, in press) where nearly identical estimates to those listed above were obtained with half as many RFLP loci. Recognizing that generalization is unwarranted based on this single example, it seems likely that stable diversity estimates will often be achieved with a sampling of as few as 50 loci.

At the time of this writing, relatively few reports of RAPD diversity have been published (Arnold et al. 1991; Van Heusden and Bachmann 1992; Wilde et al. 1992), although we expect this number to dramatically increase in the next several years. Relatively few RAPD loci were surveyed in these studies, so no conclusions are warranted regarding RAPD diversity in comparison with RFLP or isozyme diversity.

There are a number of preliminary studies involving drDNA fingerprints (Dallas 1988; Nybom 1990; Nybom and Rogstad 1990; Nybom and Schaal 1990; Nybom et al. 1990; Rogstad et al. 1991a,b; Van Houten et al. 1991), although these data do not generally allow direct comparisons with RFLP or isozyme diversity estimates. These studies have shown a disconcerting relationship between restriction enzyme used and estimate of the proportion of shared bands (D values). Nonetheless, D values have been shown to correlate with pedigree, mode of sexual reproduction, and geographic distance. Nybom (1990), for example, investigated the D values between progeny and maternal

parent, and between progeny families from six apple trees using the M13 probe. The highest D values were obtained between maternal parents and their progeny with values ranging from 0.55 to 0.85. Values dropped somewhat when D was calculated within progeny arrays (0.56–0.80) and were lowest when comparisons were made between different progeny arrays (0.23–0.53). In a sample of *Rubus* species from a single locality, Nybom and Schaal (1990) found that the sexual species *R. occidentalis* contained more multilocus genotypes than the apomictic species *R. pensilvanicus*. Finally, Nybom and Rogstad (1990) and Rogstad et al. (1991b) found that D values correlated with geographic distance in *Acer negundo* and *Asimina triloba*, respectively.

3.2. *Evolutionary process*

In addition to their utility in addressing issues of biological diversity and organismal relationships, molecular markers, especially RFLPs and RAPDs, provide a powerful set of tools for enhancing our understanding of certain evolutionary phenomena and mechanisms. We will discuss briefly how this innovation can be applied in studies of (1) genome evolution, (2) gene flow and introgression, and (3) the genetic basis of morphological evolution.

3.2.1. *Genome evolution*

The nuclear genome has played a central role in the formation of theories on plant speciation and the evolution of reproductive barriers (Grant 1981). Classical cytogenetics and biosystematics have contributed most of the information regarding chromosomal features that distinguish taxa, such as translocations, deletions, and inversions (e.g. Lewis 1962, 1966). The classical approaches are limited, however, to observation of gross structural changes in chromosomes. Moreover, they are applicable only to those taxa that can be successfully hybridized. Analysis of genome structure with molecular markers should overcome these limitations and thus offer new insights into the details of genome evolution.

A genetic linkage map is prerequisite for many studies of genome evolution. The linkage map presents the linear order of markers in their respective linkage groups and the distances between adjacent markers as a function of recombination distance. The production of linkage maps of molecular markers requires sexual reproduction and a minimum of two hybrid generations (F_1 plus F_2 or BC). Map making is straightforward for any self-compatible species since one can cross two individuals and generate a segregating F_2 population. The situation is more complex for self-incompatible species unless one is starting with parental lines that are homozygous for different polymorphisms – a common situation for many crop plants. Considerable effort may be required to achieve these conditions for self-incompatible wild populations. In the case of strictly self-incompatible, highly heterozygous plants alternative mapping strategies must be employed, such as intercrossing heterozygous parents of unknown pedigree. Enhanced methods to determine recombination frequencies

from heterozygous parents (e.g. Ritter et al. 1990) should facilitate this approach. Regardless of the starting materials, initial surveys of potential parents must be made to determine polymorphism levels and, in the case of RFLP analysis, to identify restriction enzymes that will result in segregating progeny.

Among the issues that can be addressed with detailed molecular marker maps are those that concern the manner in which plant genomes change over evolutionary time. Molecular maps provide the most powerful approach available for asking questions about chromosome evolution. In particular, molecular maps provide the resolution needed to determine the level of gene order conservation in a group of taxa, the fate of duplicated sequences and chromosome segments in aneuploids and polyploids, the level of cryptic structural differentiation (sensu Stephens 1949, 1950; Stebbins 1971) between taxa that lack apparent rearrangements (as determined from microscopy), and the genomic contribution of parental taxa to suspected hybrid species or introgressant populations.

Most of these questions require the construction of two or more linkage maps using a common set of marker loci that will each segregate in two or more mapping populations. This comparative genome mapping strategy requires that the markers employed reveal homologous polymorphisms. Because of this, RFLPs are the method of choice for comparative mapping studies, although if used cautiously and with appropriate controls, RAPDs may also provide useful comparative mapping information. Comparative mapping was pioneered by Tanksley and co-workers, who first mapped the tomato and chili pepper genomes with a set of tomato RFLP probes (Tanksley et al. 1988). More recent work has included comparative mapping of *Pisum* versus *Lens* (Weeden et al. 1992), *Sorghum* versus *Zea* (Hulbert et al. 1990; Whitkus et al. 1992), several species of *Helianthus* (Rieseberg et al. 1993), and species in *Brassica* (McGrath and Quiros 1991).

The most dramatic result that emerged from the first RFLP comparative mapping study (Tanksley et al. 1988) was that the tomato and chili pepper genomes show extensive rearrangements relative to each other, despite the fact that these members of the Solanaceae have the same haploid chromosome number ($n = 12$). Tanksley et al. (1988) estimated that a minimum of 32 chromosome breaks are necessary to account for the differences in the two genomes, indicating that the two genomes have undergone extensive rearrangement since they diverged from their common ancestor. A subsequent study, using tomato RFLP probes to map the more closely related potato genome, demonstrated that the linear order of homologous loci between tomato and potato is highly conserved, with 9 of the 12 chromosomes sharing identical locus order (Bonierbale et al. 1988). Only three rearrangements were found, each representing a paracentric inversion (Bonierbale et al. 1988).

Genomic changes that accompany aneuploidy and polyploidy are also approachable with comparative genetic mapping. This type of analysis is exemplified by the work on maize and sorghum. The genome of maize is

functionally diploid, but there is a long-standing proposal that the maize genome was derived from an $n = 5$ ancestor by allopolyploidy (Anderson 1945) or through a series of independent chromosome segment duplications (Rhoades 1951). Supporting evidence is provided by pairing between nonhomologous chromosomes in monoploid maize (Ting 1966), inheritance data indicative of duplicated loci controlling morphological traits (Rhoades 1951), and the presence of duplicated isozyme loci and linkage groups (Goodman and Stuber 1983; Wendel et al. 1989). A survey of 217 RFLP loci in maize demonstrated that its genome contains a large number of duplicated regions (Helentjaris et al. 1988). Although some chromosome pairs share a large number of duplicated loci, many in the same linear order, the data are not easily reconciled with either allopolyploidy or extensive chromosome segment duplication. Helentjaris et al. (1988) discuss a number of mechanisms that might have played a role in the generating the present genomic structure of maize, but admit that it is unclear what the relative importance of each of these mechanisms has been.

A parallel situation is seen in cultivated sorghum, a distant relative to maize but one in the same grass tribe, Andropogoneae. Again, sorghum has a functionally diploid genome, but the haploid chromosome number of 10 contrasts with a base number of 5 for the genus, suggesting that the evolutionary history of $n = 10$ sorghums includes an episode of polyploidy. Two studies have utilized maize RFLP probes to map linkage groups in sorghum (Hulbert et al. 1990; Whitkus et al. 1992). Both studies concur in demonstrating that the number of duplicated RFLP loci in sorghum is fewer than in maize, although this may be an artifact of using heterologous probes, which may have resulted in fewer hybridization signals because of the wash stringencies used. Additionally, each linkage group in sorghum appears to be composed of loci that are derived from two chromosomes of maize. Hulbert et al. (1990) interpret this result to suggest some of the loci mapped in sorghum are paralogous to those mapped in maize. Whitkus et al. (1992) hypothesize that the results reflect independent rounds of duplication in the sorghum and maize lineages, which leads to the expectation that orthologous loci need not occur in exactly the same linkages.

This complex pattern of duplication and divergence is also illustrated by comparative mapping analysis of several species in the genus *Brassica*, where both polyploidy and aneuploidy have played a role in genome evolution (U 1935; Prakash and Hinata 1980). *Brassica rapa* ($n = 10$) is considered one extreme of an aneuploid series that was initiated with an $n = 7$ ancestor (Song et al. 1990). Mapping of RFLP loci in *B. rapa* reveals the presence of numerous duplicated loci in small linkage groups in which linear order and map distances are conserved (Song et al. 1991). These data are interpreted as evidence of either internal chromosome duplications or introgression from a closely related species that contributed additional chromosome pieces. In either case, the data suggest portions of the genome that likely represent the past duplications which led to the aneuploid increase in *B. rapa*, either accompanied or followed by structural rearrangements. The results for *B. rapa* are similar to those for *B.*

oleracea ($n = 9$) in that chromosome segment duplication was associated with a fairly high degree of rearrangement, leading to the present absence of large blocks of homoeology (Slocum et al. 1990; McGrath and Quiros 1991).

A novel approach to genome mapping is seen in a study by Rieseberg et al. (1993), utilizing RAPD and isozyme loci. They suggest that the genome of *Helianthus anomalous*, a diploid species putatively derived by hybridization between *H. annuus* and *H. petiolaris*, contains large intact segments of the genomes of its two proposed parents. Although in initial stages, this study demonstrates that comparative mapping can be a powerful tool for examining speciation mechanisms.

3.2.2. Gene flow and introgression

Gene flow has long been recognized between plant populations and taxa, but its full evolutionary significance remains unclear. Although a number of evolutionary consequences of interspecific gene flow and introgression have been proposed, unambiguous empirical support for most proposed consequences is relatively scarce (reviewed in Rieseberg and Wendel 1993). Part of the difficulty has been methodological, i.e. that the available pool of molecular markers, principally isozymes during the past 20 years or so, has been relatively small. With a virtually unlimited locus number, molecular markers such as RFLPs and RAPDs represent a quantum leap forward in experimental power. Moreover, the possibility of constructing detailed linkage maps offers an as yet unexploited ability to detail the genomic extent of gene flow and introgression on a chromosome by chromosome basis. Despite these advantages of molecular markers to investigations of gene flow and introgression, relatively few have exploited this potential. This is likely to dramatically change in the next several years, as the tools continue to become more accessible to both systematists and population biologists. At present, only three published examples are particularly relevant.

The first concerns one of the classic examples of putative introgression in plants, namely, the 'Louisiana irises'. Low levels of gene flow are purported to result in populations that are morphologically intermediate to *Iris fulva* and *I. hexagona* (Riley 1938; Anderson 1949). Recently, bidirectional gene flow between the two species has been demonstrated by allozyme and ribosomal DNA analysis (Arnold et al. 1990a,b; Nason et al. 1992). In an analysis using several RAPDs, Arnold et al. (1991) detected three markers that are diagnostic for the parental species. One population that had previously been shown to consist of advanced generation segregants (Arnold et al. 1990a) contained amplification products diagnostic of both parental forms. Additionally, one allopatric population each of *I. fulva* and *I. hexagona* contained low frequency RAPD fragments that are diagnostic of the alternate species. The spatial distribution of the RAPD markers is consistent with the view that gene flow continues to occur at relatively low frequency among the two species and that introgression is both local and dispersed (Arnold et al. 1991).

The ability of molecular markers to detail the genetic structure of hybrid

zones was demonstrated by Keim et al. (1989) for a putatively introgressant population of two species of *Populus*. Diagnostic RFLP markers, identified for pure populations of *P. fremontii* and *P. angustifolia*, were used to examine the genotypic structure of individuals from a mixed population. Multilocus genotypes fell into classes representing pure *P. fremontii*, pure *P. angustifolia*, and two classes of mixed genotypes, corresponding to F₁ hybrids (heterozygous at all marker loci) and backcrosses to *P. angustifolia*. No cases were detected where progeny could be attributed to intercrossing among F₁ hybrids or crossing between F₁ plants and *P. fremontii*. These data elegantly demonstrated asymmetrical gene flow from *P. fremontii* to *P. angustifolia*.

A final example illustrates the use of RFLP analysis in studies of introgression between crop plants and their wild and/or weedy relatives. Aldrich and Doebley (1992) studied 56 accessions of wild and cultivated *Sorghum bicolor* from a wide geographic range in Africa. Their data, in conjunction with data from other sources (Duvall and Doebley 1990; Aldrich et al. 1992), support the suggestion that low levels of introgression between wild and cultivated sorghum are quite common.

3.2.3. Evolution of morphological traits

Because morphological variation forms the basis of most systems of classification, there has been a long standing desire among systematists to better understand the genetic basis of morphological evolution. Two central issues are how many genes are involved in the evolution of a new trait and what are the relative magnitudes of their effects. The neo-Darwinian school of evolution has long held that the evolution of new traits typically involves changes at many loci with individually small effects (Charlesworth et al. 1982; Coyne and Lande 1985). Nevertheless, there is little good empirical data supporting this view, especially from plant species (Gottlieb 1984; Coyne et al. 1991). Thus, it is not surprising that several authors have recently challenged the neo-Darwinian view, suggesting that evolutionary change in the morphology of plant organs often arises as the result of changes at only one or two major genes (Bachmann 1983; Hilu 1983; Gottlieb 1984).

The use of molecular markers provides a powerful means of investigating the inheritance of morphological traits. This can be done by generating a segregating F₂ population from a cross of two closely related and cross-compatible species that differ strikingly for some morphological trait. The F₂ plants are scored for their degree of expression of the trait and genotyped at a series of molecular markers distributed throughout the genome. One then examines the data for linkages between the molecular markers and the traits by either linear regression or interval mapping (Edwards et al. 1987; Lander and Botstein 1989; Knapp et al. 1990). By this procedure, one can determine both the minimal number of QTLs (quantitative trait loci) affecting a trait and the percentage of phenotypic variance that each QTL controls.

Doebley and co-workers (Doebley et al. 1990; Doebley and Stec 1991) took this approach to investigate the inheritance of several morphological traits

distinguishing maize from its near wild relative and probable progenitor, teosinte. The results showed that some traits have patterns of inheritance that appear to involve a single major locus plus several loci of small effect. Other traits showed more multigenic inheritance, being controlled by four or five loci of roughly equal effects. These authors suggested that evolution is 'opportunistic', i.e. involving genes with all ranges of effects from small to large.

Paterson et al. (1991) employed molecular markers to investigate the inheritance of fruit mass in segregating F_2 , F_3 and backcross populations of tomato and related wild species. They detected 11 QTLs affecting this trait that individually controlled between 4.0 and 42.0% of the phenotypic variance. The authors raised the 'intriguing possibility' that a small subset of loci control much of the variance involved in the evolution of fruit mass. In this sense, their data agree with that of Doebley and co-workers who also found relatively few QTLs controlling most individual traits.

Fatokun et al. (1992) investigated the inheritance of seed weight in two species of cultivated legumes, cowpea and mungbean. Since both species belong to the genus *Vigna*, it was possible to use a common set of molecular markers to analyze one F_2 population derived from a cross of wild and cultivated cowpea and another from a cross of wild and cultivated mungbean. These authors detected two QTLs for seed weight in the cowpea population that controlled 36% and 32% of the variance. They detected four QTLs in the mungbean population, the largest one controlling 32% of the variance. Remarkably, the QTL of largest effect in both species was located near the same RFLP loci, suggesting that mutations of large effect were independently selected at the same locus during the evolution of these two crops. This provides further evidence that a limited set of loci may control much of the variance upon which selection acts when a trait evolves and that QTLs with large effects may be a common feature of plant evolution.

While the studies just reviewed provide compelling evidence that QTLs with large effects have been important in the evolution of maize, tomato, cowpea and mungbean, these are all crop species. It will be of interest to apply QTL-mapping to natural species and determine if evolution under natural selection operates in a similar mode.

4. Conclusion

Molecular markers have proven to be a useful tool in several areas of plant biology and are just beginning to be exploited in plant systematics. They are ideally suited to systematic questions that require a combination of extensive genomic coverage and a large number of variable loci. These features offer greater resolution for a number of evolutionary questions that are less approachable with isozyme analysis, gene sequencing or restriction site analysis of chloroplast DNA. In particular, molecular markers provide a heretofore unparalleled ability to address questions of genome evolution, dissect the

genetic basis underlying morphological evolution, describe the multilocus genotypic structure of populations, and detail the genomic distribution and pattern of inter-taxon gene flow and introgression. As with many technological developments, molecular markers provide new avenues of research as well as novel approaches to long-standing problems. Their continued use will undoubtedly enhance our understanding in many areas of plant systematics.

5. References

- Aldrich, P R and Doebley, J (1992) Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor*. *Theor Appl Genet* 85 293–302
- Aldrich, P R, Doebley, J, Schertz, K F and Stec, A (1992) Patterns of allozyme variation in cultivated and wild *Sorghum bicolor*. *Theor Appl Genet* 85 451–460
- Anderson, E (1945) What is *Zea mays*? A report of progress. *Chron Bot* 9 88–92
- Anderson, E (1949) *Introgressive Hybridization*. John Wiley, New York
- Arnold, M L, Bennett, B D and Zimmer, E A (1990a) Natural hybridization between *Iris fulva* and *I hexagona*: pattern of ribosomal DNA variation. *Evolution* 44 1512–1521
- Arnold, M L, Hamrick, J L and Bennett, B D (1990b) Allozyme variation in Louisiana irises: a test for introgression and hybrid speciation. *Heredity* 65 297–306
- Arnold, M L, Buckner, C M and Robinson, J J (1991) Pollen-mediated introgression and hybrid speciation in Louisiana irises. *Proc Natl Acad Sci U S A* 88 1398–1402
- Bachmann, K (1983) Evolutionary genetics and the genetic control of morphogenesis in flowering plants. *Evol Biol* 16 157–208
- Ball, R M Jr, Neigel, J E and Avise, J C (1990) Gene genealogies within the organismal pedigrees of random mating populations. *Evolution* 44 1109–1119
- Baverstock, P R and Moritz, C (1990) Sampling design. In: D M Hillis and C Moritz (eds), *Molecular Systematics*, pp 13–24. Sinauer Associates, Inc., Sunderland, MA
- Beckmann, J S and Soller, M (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. *Theor Appl Genet* 67 35–43
- Bonierbale, M W, Plaisted, R L and Tanksley, S D (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* 120 1095–1103
- Brubaker, C L and Wendel, J F (1994) Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*, Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). *Am J Bot* (in press)
- Brummer, E C, Kochert, G and Bouton, J H (1991) RFLP variation in diploid and tetraploid alfalfa. *Theor Appl Genet* 83 89–96
- Chang, C, Bowman, J L, DeJohn, A W, Lander, E S and Meyerowitz, E M (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 85 6856–6860
- Charlesworth, B, Lande, R and Slatkin, M (1982) A neo-Darwinian commentary on macroevolution. *Evolution* 36 474–498
- Chase, C D, Ortega, V M and Vallejos, C E (1991) DNA restriction fragment length polymorphisms correlate with isozyme diversity in *Phaseolus vulgaris* L. *Theor Appl Genet* 81 806–811
- Coyne, J A and Lande, R (1985) The genetic basis of species differences in plants. *Am Nat* 126 141–145
- Coyne, J A, Rux, J and David, J R (1991) Genetics of morphological differences and hybrid sterility between *Drosophila sechellia* and its relatives. *Genet Res, Camb* 57 113–122
- Crawford, D J (1983) Phylogenetic and systematic inferences from electrophoretic studies. In: S D Tanksley and T J Orton (eds), *Isozymes in Plant Genetics and Breeding, Part A*, pp 257–287. Elsevier, Amsterdam

- Crawford, D J (1990) Plant Molecular Systematics John Wiley and Sons, Inc , New York
- Dallas, J F (1988) Detection of DNA 'fingerprints' of cultivated rice by hybridization with a human minisatellite DNA probe Proc Natl Acad Sci U S A 85 6831-6835
- Doebley, J and Stec, A (1991) Genetic analysis of the morphological differences between maize and teosinte Genetics 129 285-295
- Doebley, J , Stec, A , Wendel, J and Edwards, M (1990) Genetic and morphological analysis of a maize-teosinte F₂ population implications for the origin of maize Proc Natl Acad Sci U S A 87 9888-9892
- Doyle, J J (1992) Gene trees and species trees molecular systematics as one-character taxonomy Syst Bot 17 144-163
- Duvall, M R and Doebley, J F (1990) Restriction site variation in the chloroplast genome of *Sorghum* (Poaceae) Syst Bot 15 472-480
- Edwards, M D , Stuber, C W and Wendel, J F (1987) Molecular-marker-facilitated investigations of quantitative-trait loci in maize I Numbers, genomic distribution and types of gene action Genetics 116 113-125
- Epplen, J T (1988) On simple repeated GATA/GACA sequences in animal genomes a critical reappraisal J Hered 79 409-417
- Evola, S V , Burr, F A and Burr, B (1986) The suitability of restriction fragment length polymorphisms as genetic markers in maize Theor Appl Genet 71 765-771
- Fatokun, C A , Menancio-Hautea, D I , Danesh, D and Young, N D (1992) Evidence for orthologous seed weight genes in cowpea and mungbean based on RFLP mapping Genetics 132 841-846
- Felsenstein, J (1988) Phylogenies from molecular sequences inference and reliability Ann Rev Genet 22 521-565
- Figdore, S S , Kennard, W D , Song, K M , Slocum, M K and Osborn, T C (1988) Assessment of the degree of restriction fragment length polymorphism in *Brassica* Theor Appl Genet 75 833-840
- Flavell, R (1986) Repetitive DNA and chromosome evolution in plants Phil Trans R Soc London Ser B 312 227-242
- Georges, M , Lathrop, M , Hilbert, P , Marcotte, A , Schwers, A , Swillens, S , Vassart, G and Hanset, R (1990) On the use of DNA fingerprints for linkage studies in cattle Genomics 6 461-474
- Gepts, P (1993) The use of molecular and biochemical markers in crop evolution studies Evol Biol 27 51-94
- Goodman, M M and Stuber, G W (1983) Maize In S D Tanksley and T J Orton (eds), Isozymes in Plant Genetics and Breeding, Part B, pp 1-33 Elsevier, Amsterdam
- Gottlieb, L D (1981) Electrophoretic evidence and plant populations Prog Phytochem 7 1-46
- Gottlieb, L D (1984) Genetics and morphological evolution in plants Am Nat 123 681-709
- Graner, A , Jahoor, A , Schondelmaier, J , Siedler, H , Pillen, K , Fischbeck, G , Wenzel, G and Herrmann, R G (1991) Construction of an RFLP map of barley Theor Appl Genet 83 250-256
- Grant, V (1981) Plant Speciation, 2nd ed Columbia University Press, New York
- Guo, M , Lightfoot, D A , Mok, M C and Mok, D W S (1991) Analyses of *Phaseolus vulgaris* L and *P. coccineus* Lam hybrids by RFLP preferential transmission of *P. vulgaris* alleles Theor Appl Genet 81 703-709
- Halward, T M , Stalker, H T , LaRue, E A and Kochert, G (1991) Genetic variation detectable with molecular markers among unadapted germ-plasm resources of cultivated peanut and related wild species Genome 34 1013-1020
- Hamby, R K and Zimmer, E A (1992) Ribosomal RNA as a phylogenetic tool in plant systematics In P S Soltis, D E Soltis and J J Doyle (eds), Molecular Systematics of Plants, pp 50-91 Chapman and Hall, New York
- Hamrick, J L and Godt, M J W (1990) Allozyme diversity in plant species In A H D Brown, M T Clegg, A L Kahler and B S Weir (eds), Plant Population Genetics, Breeding, and Genetic Resources, pp 43-63 Sinauer Associates, Inc , Sunderland, MA
- Hartl, D L and Clark, A G (1989) Principles of Population Genetics, 2nd ed Sinauer Associates, Inc , Sunderland, MA

- Helentjaris, T, King, G, Slocum, M, Siedenstrang, C and Wegman, S (1985) Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding *Plant Mol Biol* 5 109–118
- Helentjaris, T, Slocum, M, Wright, S, Schaefer, A and Nienhuis, J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms *Theor Appl Genet* 72 761–769
- Helentjaris, T, Weber, D and Wright, S (1988) Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms *Genetics* 118 353–363
- Heun, M, Kennedy, A E, Anderson, J A, Lapitan, N L V, Sorrells, M E and Tanksley, S D (1991) Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*) *Genome* 34 437–447
- Hillis, D M and Moritz, C (eds) (1990) *Molecular Systematics* Sinauer Associates, Inc, Sunderland, MA
- Hilu, K W (1983) The role of single-gene mutations in the evolution of flowering plants *Evol Biol* 26 97–128
- Hosaka, K, Kianian, S F, McGrath, J M and Quiros, C F (1990) Development and chromosomal localization of genome-specific DNA markers of *Brassica* and the evolution of amphidiploids and $n = 9$ diploid species *Genome* 33 131–142
- Hulbert, S H, Richter, T E, Axtell, J D and Bennetzen, J L (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes *Proc Natl Acad Sci U S A* 87 4251–4255
- Jeffreys, A J, Wilson, V and Thein, S L (1985a) Hypervariable ‘minisatellite’ regions in human DNA *Nature* 314 67–73
- Jeffreys, A J, Wilson, V and Thein, S L (1985b) Individual-specific ‘fingerprints’ of human DNA *Nature* 316 76–79
- Jeffreys, A J, Wilson, V, Kelly, R, Taylor, B A and Bulfield, G (1987) Mouse DNA ‘fingerprints’ analysis of chromosome localization and germ-line stability of hypervariable loci in recombinant inbred strains *Nucl Acids Res* 15 2823–2836
- Jeffreys, A J, Royle, N J, Wilson, V and Wong, Z (1988) Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA *Nature* 332 278–281
- Jena, K K and Kochert, G (1991) Restriction fragment length polymorphism analysis of CCDD genome species of the genus *Oryza* L *Plant Mol Biol* 16 831–839
- Jones, C W and Kafatos, F C (1982) Accepted mutations in a gene family evolutionary diversification of duplicated DNA *J Mol Evol* 19 87–103
- Keim, P, Paige, K N, Whitham, T G and Lark, K G (1989) Genetic analysis of an interspecific hybrid swarm of *Populus* occurrence of unidirectional introgression *Genetics* 123 557–565
- Kesseli, R, Ochoa, O and Michelmore, R (1991) Variation at RFLP loci in *Lactuca* subsp and origin of cultivated lettuce (*L sativa*) *Genome* 34 430–436
- Kim, K -J, Jansen, R K, Wallace, R S, Michaels, H J and Parker, J D (1992) Phylogenetic implications of *rbcL* sequence variation in the Asteraceae *Ann Missouri Bot Gard* 79 428–445
- Kochert, G, Halward, T, Branch, W D and Simpson, C E (1991) RFLP variability in peanut (*Arachis hypogaea* L) cultivars and wild species *Theor Appl Genet* 81 565–570
- Knapp, S J, Bridges, W C Jr and Birkes, D (1990) Mapping quantitative trait loci using molecular marker linkage maps *Theor Appl Genet* 79 583–592
- Lander, E S and Botstein, D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps *Genetics* 121 185–199
- Landry, B S and Michelmore, R W (1987) Methods and applications of restriction fragment length polymorphism analysis to plants In G Bruening, J Harada, T Kosuge and A Hollaender (eds), *Tailoring Genes for Crop Improvement*, pp 25–44 Plenum Publishing Corporation, New York
- Landry, B S, Kesseli, R, Leung, H and Michelmore, R W (1987) Comparison of restriction endonucleases and sources of probes for their efficiency in detecting restriction fragment length polymorphisms in lettuce (*Lactuca sativa* L) *Theor Appl Genet* 74 646–653
- Lewin, B (1990) *Genes IV* Oxford University Press, Oxford

- Lewis, H (1962) Catastrophic selection as a factor in speciation *Evolution* 16 257–271
- Lewis, H (1966) Speciation in flowering plants *Science* 152 167–172
- Liu, Y G , Mori, N and Tsuenewaki, K (1990) Restriction fragment length polymorphism (RFLP) analysis of wheat I Genomic DNA library construction and RFLP analysis in common wheat *Jpn J Genet* 65 367–380
- Loveless, M D and Hamrick, J L (1984) Ecological determinants of genetic structure in plant populations *Ann Rev Ecol Syst* 15 65–95
- Lynch, M (1988) Estimation of relatedness by DNA fingerprinting *Mol Biol Evol* 5 584–599
- Lynch, M (1990) The similarity index and DNA fingerprinting *Mol Biol Evol* 7 478–484
- Lynch, M (1991) Analysis of population genetic structure by DNA fingerprinting In T Burke, G Dolf, A J Jeffreys and R Wolff (eds), *DNA Fingerprinting Approaches and Applications*, pp 113–126 Birkhauser Verlag, Basel, Switzerland
- McClellan, P E and Hanson, M R (1986) Mitochondrial DNA sequence divergence among *Lycopersicon* and related *Solanum* species *Genetics* 112 649–667
- McCouch, S R , Kochert, G , Yu, Z H , Wang, Z Y , Khush, G S , Coffman, W R and Tanksley, S D (1988) Molecular mapping of rice chromosomes *Theor Appl Genet* 76 815–829
- McGrath, J M and Quiros, C F (1991) Inheritance of isozyme and RFLP markers in *Brassica campestris* and comparison with *B oleracea* *Theor Appl Genet* 82 668–673
- McGrath, J M and Quiros, C F (1992) Genetic diversity at isozyme and RFLP loci in *Brassica campestris* as related to crop type and geographical origin *Theor Appl Genet* 83 783–790
- Miller, J C and Tanksley, S D (1990a) Effect of different restriction enzymes, probe source, and probe length on detecting restriction fragment length polymorphism in tomato *Theor Appl Genet* 80 385–389
- Miller, J C and Tanksley, S D (1990b) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon* *Theor Appl Genet* 80 437–448
- Miyamoto, M M and Cracraft, J (eds) (1991) *Phylogenetic Analysis of DNA Sequences* Oxford University Press, New York
- Morden, C W , Doebley, J F and Schertz, K F (1989) Allozyme variation in old world races of *Sorghum bicolor* (Poaceae) *Am J Bot* 76 247–255
- Morden, C W , Doebley, J F and Schertz, K F (1990) Allozyme variation among the spontaneous species of *Sorghum* section *Sorghum* (Poaceae) *Theor Appl Genet* 80 296–304
- Nason, J D , Ellstrand, N C and Arnold, M L (1992) Patterns of hybridization and introgression in populations of oaks, manzanitas, and irises *Am J Bot* 79 101–111
- Nei, M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals *Genetics* 89 583–590
- Nei, M (1987) *Molecular Evolutionary Genetics* Columbia University Press, New York
- Nei, M , Tajima, F and Tateno, Y (1983) Accuracy of estimated phylogenetic trees from molecular data II Gene frequency data *J Mol Evol* 19 153–170
- Neuhausen, S L (1992) Evaluation of restriction fragment length polymorphism in *Cucumis melo* *Theor Appl Genet* 83 379–384
- Nybom, H (1990) Genetic variation in ornamental apple trees and their seedlings (*Malus*, Rosaceae) revealed by DNA ‘fingerprinting’ with the M13 repeat probe *Hereditas* 113 17–28
- Nybom, H and Rogstad, S H (1990) DNA ‘fingerprints’ detect genetic variation in *Acer negundo* (Aceraceae) *Plant Syst Evol* 173 49–56
- Nybom, H and Schaal, B A (1990) DNA ‘fingerprints’ reveal genotypic distributions in natural populations of blackberries and raspberries (*Rubus*, Rosaceae) *Am J Bot* 77 883–888
- Nybom, H , Schaal, B A and Rogstad, S H (1989) DNA fingerprints can distinguish cultivars of blackberries and raspberries *Acta Hort* 262 305–310
- Nybom, H , Rogstad, S H and Schaal, B A (1990) Genetic variation detected by use of the M13 ‘DNA fingerprint’ probe in *Malus Prunus*, and *Rubus* (Rosaceae) *Theor Appl Genet* 79 153–156
- Palmer, J D and Zamir, D (1982) Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon* *Proc Natl Acad Sci U S A* 79 5006–5010
- Palmer, J D , Jansen, R K , Michaels, H J , Chase, M W and Manhart, J R (1988) Chloroplast

- DNA variation and plant phylogeny *Ann Missouri Bot Gard* 75 1180–1206
- Paran, I, Kesseli, R and Michelmore, R (1991) Identification of restriction fragment length polymorphism and random amplified polymorphic DNA markers linked to downy mildew resistance genes in lettuce, using near-isogenic lines *Genome* 34 1021–1027
- Paterson, A H, Damon, S, Hewitt, J D, Zamir, D, Rabinowitch, H D, Lincoln, S E, Lander, E S and Tanksley, S D (1991) Mendelian factors underlying quantitative traits in tomato comparison across species, generations and environments *Genetics* 127 181–197
- Prakash, S and Hinata, K (1980) Taxonomy, cytogenetics, and origin of crop *Brassica*, a review *Opera Bot* 55 1–57
- Rhoades, M M (1951) Duplicated genes in maize *Am Nat* 85 105–110
- Riley, H P (1938) A character analysis of colonies of *Iris fulva* and *Iris hexagona* var *giganticaerulea* and natural hybrids *Am J Bot* 25 727–738
- Rieseberg, L H and Soltis, D E (1991) Phylogenetic consequences of cytoplasmic gene flow in plants *Evol Trends Plants* 5 65–84
- Rieseberg, L H and Wendel, J F (1993) Introgression and its consequences In R Harrison (ed), *Hybrid Zones and the Evolutionary Process*, pp 70–109 Oxford University Press
- Rieseberg, L H and Brunfeldt, S J (1992) Molecular evidence and plant introgression In P S Soltis, D E Soltis and J J Doyle (eds), *Molecular Systematics of Plants*, pp 151–201 Chapman and Hall, New York
- Rieseberg, L H, Choi, H, Chan, R and Spore, C D (1993) Genomic map of a diploid hybrid species *Heredity* 70 285–293
- Ritter, E, Gebhardt, G and Salamini, F (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents *Genetics* 125 645–654
- Rogstad, S H, Patton, J C and Schaal, B A (1988) A human minisatellite probe reveals RFLPs among individuals of two angiosperms *Nucl Acids Res* 16 11378
- Rogstad, S H, Nybom, H and Schaal, B A (1991a) The tetrapod ‘DNA fingerprinting’ M13 repeat probe reveals genetic diversity and clonal growth in quaking aspen (*Populus tremuloides*, Salicaceae) *Plant Syst Evol* 175 115–123
- Rogstad, S H, Wolff, K and Schaal, B A (1991b) Geographical variation in *Asimina triloba* Dunal (Annonaceae) revealed by the M13 ‘DNA fingerprinting’ probe *Am J Bot* 78 1391–1396
- Ryskov, A P, Jincharadze, A G, Prosnjak, M I, Ivanov, P L and Limborska, S A (1988) M13 phage DNA as a universal marker for DNA fingerprinting of animals, plants and microorganisms *FEBS Lett* 233 388–392
- Sambrook, J, Fritsch, E F and Maniatis, T (1989) *Molecular Cloning A Laboratory Manual*, 2nd ed Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Shattuck-Eidens, D M, Bell, R N, Neuhausen, S L and Helentjaris, T (1990) DNA sequence variation within maize and melon observations from polymerase chain reaction amplification and direct sequencing *Genetics* 126 207–217
- Slocum, M K, Figdore, S S, Kennard, W C, Suzuki, J Y and Osborn, T C (1990) Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea* *Theor Appl Genet* 80 57–64
- Smith, G P (1976) Evolution of repeated DNA sequences by unequal crossover *Science* 191 528–535
- Smith, J S C and Smith, O S (1991) Restriction fragment length polymorphisms can differentiate among U S maize hybrids *Crop Sci* 31 893–899
- Soltis, D E and Soltis, P S (eds) (1989) *Isozymes in Plant Biology* Dioscorides Press, Portland, OR
- Soltis, D E, Soltis, P S and Milligan, B G (1992a) Intraspecific chloroplast DNA variation systematic and phylogenetic implications In P S Soltis, D E Soltis and J J Doyle (eds), *Molecular Systematics of Plants*, pp 117–150 Chapman and Hall, New York
- Soltis, P S, Soltis, D E and Doyle, J J (eds) (1992b) *Molecular Systematics of Plants* Chapman and Hall, New York
- Song, K M, Osborn, T C and Williams, P H (1988) *Brassica* taxonomy based on nuclear

- restriction fragment length polymorphisms (RFLPs) 1 Genome evolution of diploid and amphidiploid species *Theor Appl Genet* 75 784–794
- Song, K M , Osborn, T C and Williams, P H (1990) *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs) 3 Genome relationships in *Brassica* and related genera and the origin of *B oleracea* and *B rapa* (syn *campestris*) *Theor Appl Genet* 79 497–506
- Song, K M , Suzuki, J Y , Slocum, M K , Williams, P H and Osborn, T C (1991) A linkage map of *Brassica rapa* (syn *campestris*) based on restriction fragment length polymorphism loci *Theor Appl Genet* 82 296–304
- Stebbins, G L (1971) *Chromosomal Evolution in Higher Plants* Edward Arnold, London
- Stephens, S G (1949) The cytogenetics of speciation in *Gossypium* I Selective elimination of the donor parent genotype in interspecific backcrosses *Genetics* 34 627–637
- Stephens, S G (1950) The internal mechanism of speciation in *Gossypium* *Bot Rev* 16 115–149
- Swofford, D L and Olsen, G J (1990) Phylogeny reconstruction. In D M Hillis and C Moritz (eds), *Molecular Systematics*, pp 411–501 Sinauer Associates, Inc , Sunderland, MA
- Tanksley, S D , Bernatzky, R , Lapitan, N L and Prince, J P (1988) Conservation of gene repertoire but not gene order in pepper and tomato *Proc Natl Acad Sci U S A* 85 6419–6423
- Tautz, D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers *Nucl Acids Res* 17 6463–6471
- Tautz, D and Renz, M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes *Nucl Acids Res* 17 6463–6471
- Ting, Y C (1966) Duplications and meiotic behavior of the chromosomes in haploid maize (*Zea mays* L) *Cytologia* 31 324–329
- U, N (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B napus* and peculiar mode of fertilization *Jpn J Bot* 7 389–452
- Van de Ven, M , Powell, W , Ramsay, G and Waugh, R (1990) Restriction fragment length polymorphisms as genetic markers in *Vicia* *Heredity* 65 329–342
- Van Heusden, A W and Bachmann, K (1992) Genotype relationships in *Microseris elegans* (Asteraceae, Lactuceae) revealed by DNA amplification from arbitrary primers (RAPDs) *Plant Syst Evol* 179 221–233
- Van Houten, W H J , van Heusden, A W , Rouppe, J , Voort, V D , Rajmanna, L and Bachmann, K (1991) Hypervariable DNA fingerprint loci in *Microseris pygmaea* (Asteraceae, Lactuceae) *Bot Acta* 104 252–265
- Vassart, G , Georges, M , Monsieur, R , Brocas, H , Lequarre, A S and Christophe, D (1987) A sequence in M13 phage detects hypervariable minisatellite in human and animal DNA *Science* 235 683–684
- Walsh, J B (1987) Persistence of tandem arrays implications for satellite and simple-sequence DNA *Genetics* 115 553–567
- Weeden, N F , Muehlbauer, F J and Ladizinsky, J (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps *J Hered* 83 123–129
- Weir, B S (1990a) *Genetic Data Analysis* Sinauer Associates, Inc , Sunderland, MA
- Weir, B S (1990b) Intraspecific differentiation. In D M Hillis and C Moritz (eds), *Molecular Systematics*, pp 373–410 Sinauer Associates, Inc Sunderland, MA
- Wendel, J F , Stuber, C W , Goodman, M M and Beckett, J B (1989) Duplicated plastid and triplicated cytosolic isozymes of triosephosphate isomerase in maize (*Zea mays* L) *J Hered* 80 218–228
- Wilde, J , Waugh, R and Powell, W (1992) Genetic fingerprinting of *Theobroma* clones using random amplified polymorphic DNA markers *Theor Appl Genet* 83 871–877
- Williams, J G K , Kubelik, A R , Livak, K J , Rafalski, J A and Tingey, S V (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers *Nucl Acids Res* 18 6531–6535
- Whitkus, R , Doebley, J and Lee, M (1992) Comparative genome mapping of sorghum and maize *Genetics* 132 1119–1130
- Zimmerman, P A , Lang-Unnasch, N and Cullis, C A (1989) Polymorphic regions in plant genomes detected by an M13 probe *Genome* 32 824–828

7. Introduction: molecular marker maps of major crop species

RONALD L. PHILLIPS¹ and INDRA K. VASIL²

¹ *Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108-6026, U.S.A.;* ² *Laboratory of Plant Cell and Molecular Biology, University of Florida, Gainesville, FL 32611-0690, U.S.A.*

Until very recently genetic maps of higher plants were based almost entirely on morphological and biochemical traits. These maps are rapidly being replaced and/or supplemented with DNA-based marker maps formulated on the use of powerful new molecular techniques.

The new high precision maps can be developed with comparative ease and rapidity. They have a much higher density of markers which allows revelation of more and more restricted segments of the genome. One of the many revolutionary aspects of this technology is that linkage between molecular markers and traits of interest often can be detected in a single cross. The ability to hybridize probe after probe to the DNA of the same individuals of a segregating population allows one to pursue the analysis until linkage becomes evident. With morphological and biochemical markers used previously, a separate cross was required to test linkage with each new marker. It was seldom that more than three markers could be tested for linkage with the trait of interest in a single cross because of viability problems. With the new techniques described in this volume, a new gene could be placed on the linkage map within a few days instead of the much longer time required with the previous techniques.

Recombinant inbred lines and other means to immortalize segregating lines derived from a cross coupled with appropriate software programs facilitate the building of linkage maps. Other approaches also can streamline the detection of linkage, such as the use of backcross derived lines, bulked segregant analysis, or aneuploids. Linkage maps of the future will include the DNA-based markers as well as many qualitative and quantitative trait loci. Such maps will be extended over time to provide an impressive and useful genic array displayed across the entire genome. Maps including cDNAs and other cloned genes will provide additional dimensions for determining expressed gene relationships.

The genetic dissection of quantitative traits is made feasible by DNA-based markers which have no phenotypic manifestations. Genomic regions can be recognized which control large portions of the phenotypic variation of important traits. The pyramiding of genes influencing a single trait will lead to many new insights and useful strains.

As new technologies develop that allow the automation of marker analysis, the applications in diagnostics, breeding, proprietary protection, and cloning of genes will no doubt increase. The ability to analyze large numbers of individuals at modest cost will enhance the attractiveness of molecular markers in these fields and lead to their routine use. Map-based cloning requires a high density marker map around the targeted region. The increasing interest in plant genome analysis is leading to even more saturated maps useful in the isolation of genes. This will be helpful not only in marker assisted breeding, but also in introducing desirable genes into crops of interest by genetic transformation.

In the following chapters, a group of leading researchers provide the latest version of DNA-based marker maps for a variety of important crops. These maps illustrate the state of the art today. The progress made during the past five years has been truly phenomenal.

8. Molecular maps of alfalfa

E.C. BRUMMER^{1,*}, C.S. ECHT², T.J. McCOY², K.K. KIDWELL³, T.C. OSBORN³, G.B. KISS⁴, G. CSANADI⁴, K. KALMAN⁴, J. GYÖRGYÉY⁴, L. ÖKRÉSZ⁴, A.E. RACZKEVY⁴, J.H. BOUTON¹ and G. KOCHERT¹

¹Departments of Agronomy (ECB, JHB) and Botany (GK), University of Georgia, Athens, GA 30602, U S A , ²Department of Plant and Soil Science, Montana State University, Bozeman, MT 59717-0312, U S A , ³Department of Agronomy, University of Wisconsin, Madison, WI 53706, U S A , ⁴Institute of Genetics, Biological Research Center, Hungarian Academy of Sciences, Szeged, P O Box 521, H-6701, Hungary, * Present address Department of Agronomy, Iowa State University, Ames, IA 50011, U S A

Contents

1 Introduction	144	8 Application of molecular markers to alfalfa breeding and cultivar development	152
2 Taxonomy and cytogenetics	145	9 Problems of molecular marker analysis in alfalfa	156
3 Germplasm and sources	146	10 Acknowledgements	156
4 Variability	146	11 References	157
5 Mapping – University of Georgia	146		
6 Mapping – Montana State University and University of Wisconsin	150		
7 Mapping – Hungarian Academy of Sciences, Szeged	152		

1. Introduction

Alfalfa, *Medicago sativa* L., is a highly productive forage species, grown throughout the temperate regions of the world for its premium quality hay and pasturage. Alfalfa originated near the Caspian Sea in northern Iran and northeastern Turkey; its cultivation spread throughout the Mediterranean region and into Germany, France, and China by the time of the Roman Empire (Bolton 1962). Today, alfalfa is raised on all continents and is currently cultivated on more than 32 million hectares worldwide (Michaud et al. 1988). The genus *Medicago* is highly diverse, including not only perennial, cultivated alfalfa but also several annual forage species and many species with little apparent agronomic value but having potentially vital germplasm for further improvement of the cultivated species.

Alfalfa produces high dry matter yields with high crude protein and total digestible nutrients. Yields as high as 24 Mg/ha have been attained under nonirrigated conditions (Barnes et al. 1988). Alfalfa is able to fix atmospheric nitrogen which helps improve the soil for future crops while concurrently eliminating the need for inorganic N fertilizer. In addition to these direct agronomic benefits, alfalfa is a useful model system for the study of forage systems and autotetraploid genetic theory (Barnes et al. 1988).

Nevertheless, alfalfa is not a perfect crop and further improvement is necessary. Cultivar development programs attempt to incorporate multiple pest resistances, increase yields and quality, and increase N₂ fixation. Other objectives such as acid-soil tolerance, grazing tolerance, and bloat resistance are also worthy of consideration.

2. Taxonomy and cytogenetics

Several excellent discussions of the taxonomy of alfalfa have been published (Lesins and Gillies 1972; Lesins and Lesins 1979; Quiros and Bauchan 1988). At least fifty-six species are currently recognized within the genus *Medicago* (McCoy and Bingham 1988). Several species described in earlier studies (e.g. Lesins and Gillies 1972; Quiros 1983) have been reclassified as subspecies of *M. sativa*. The basic chromosome number of most *Medicago* species, including cultivated alfalfa, is $x = 8$ although several species have $x = 7$, apparently as a result of chromosomal rearrangements (Lesins and Lesins 1979). Diploid, tetraploid, and hexaploid species are known. Both annual and perennial species are present within the genus: the annuals being autogamous and perennials allogamous. Cultivated alfalfa is perennial and autotetraploid ($2n = 4x = 32$, Stanford 1951; Quiros 1982).

Medicago sativa is actually a complex of related diploid and tetraploid subspecies which are interfertile and have the same karyotype (Quiros and Bauchan 1988). Included within the complex are four subspecies: *Medicago sativa* subsp. *falcata* ($2x$ and $4x$), subsp. *sativa* ($4x$), subsp. *coerulea* ($2x$), and subsp. *glutinosa* ($4x$), as well as a variety of hybrid subspecies. *M. sativa* subsp. *coerulea* is the diploid form of subsp. *sativa* (Quiros and Bauchan 1988). All subspecies have purple flowers except yellow flowered subsp. *falcata*. Cultivated alfalfa consists predominantly of subsp. *sativa* germplasm, but other subspecies, particularly subsp. *falcata*, have been introgressed. Endemic to colder regions, subsp. *falcata* has provided winterhardiness to cultivated alfalfa, though it has also contributed slower regrowth and a more prostrate growth habit than pure subsp. *sativa* (Lesins and Lesins 1979).

Attempts have been made to hybridize *M. sativa* germplasm with other *Medicago* species (Lesins and Lesins 1979; McCoy and Bingham 1988). Twelve perennial species have been successfully hybridized with *M. sativa* although some of the hybridizations require the use of ovule-embryo culture in order to recover the interspecific progeny (McCoy and Smith 1986). A cross of *M. sativa* \times *M. scutellata* Mill. is the only example of hybridization between annual and perennial species that has been reported (Sangduen et al. 1982). Further studies are needed to determine the extent of hybridization among other *Medicago* species.

3. Germplasms and sources

The major source of *Medicago* germplasm for research purposes in the United States is the USDA-ARS National Plant Germplasm System. Plant introductions are stored and catalogued at Pullman, Washington. Germplasm for all species has been collected from many sites worldwide and is available for research use.

Genetic studies in alfalfa have been enhanced by the ease of extracting haploids from cultivated alfalfa (Bingham 1971). Cultivated Alfalfa at the Diploid Level (CADL) is an important germplasm developed via haploidization (Bingham and McCoy 1979). Haploids ($n = 2x = 16$) were derived from tetraploid cultivated alfalfa and crossed with diploid *M. sativa* subsp. *falcata* accessions. The progeny were then repeatedly backcrossed to the cultivated parental haploids or intercrossed among themselves. The current population is essentially entirely cultivated germplasm and is reproductively stable at the diploid level. W2xiso and W4xiso, two germplasm populations developed from CADL, are useful in studies comparing populations with the same genetic constitution at two ploidy levels (Bingham 1991).

4. Variability

Extensive variability exists for many traits of alfalfa. This variability has enabled breeders to develop cultivars with a wide variety of pest resistances as well as high yield. More extensive variation is found at the species level, with many traits (e.g. pod shape) varying significantly among species (Lesins and Lesins 1979). Isozyme variability, or the existence of multiple alleles, has also been demonstrated for alfalfa (Quiros 1982, 1983). In a study of DNA polymorphism detected by RFLPs, Brummer et al. (1991) showed that large amounts of variability were present, even among plants of the same population, at both the diploid and tetraploid levels. With only nineteen cDNA clones, they differentiated every plant in the study. The high level of variation is not unexpected since alfalfa is an outcrossing species with very low self-fertility. The existence of such high levels of variation at the DNA level indicates that further studies utilizing molecular markers should be pursued.

5. Mapping – University of Georgia

Two diploid plants were selected based on a previous study of RFLP variation (Brummer et al. 1991) and used as parents for a molecular mapping project. These plants are within the *M. sativa* complex and should not contain any chromosomal rearrangements which may have resulted from a wider cross. Further, the amount of variability present within the germplasm is sufficient for a mapping study. One of the parents (3W) is derived from the W2x-iso

germplasm. The other parent (440501-2) is derived from *M. sativa* subsp. *coerulea* plant introduction #440501. These parents were crossed using hand emasculation to produce an F₁ hybrid. Selfed seed of the hybrid was used to produce a segregating F₂ population of 86 individuals which is currently being used to map the alfalfa genome.

The experimental protocols for DNA extraction, cDNA library construction, agarose electrophoresis, hybridization and RFLP analysis have been reported (Brummer et al. 1991). We have tried several DNA extraction procedures (Murray and Thompson 1980; Dellaporta et al. 1983; Apuya et al. 1988; Shure et al. 1983) and have found that the procedure of Saghai-Marooft et al. (1984) using freeze-dried tissue gives the cleanest DNA and the most consistent quantities. The cDNA probes used in mapping were isolated from the bacteriophage Lambda Zap II vector (Stratagene) using PCR (Oste 1988). The clones were initially screened with nick-translated total genomic DNA and chloroplast DNA (gift from Dr. Glenn Galau, University of Georgia) to eliminate highly repeated or chloroplast sequences, respectively. The remaining clones were then screened on filters containing DNA from the two parents and the F₁ hybrid digested with *EcoRI*, *HindIII*, and *EcoRV*. Probes showing simple patterns of codominant segregation were then screened on filters with the parents, F₁, and F₂ individuals' DNA restricted with the appropriate enzyme. Clone nomenclature is given in Table 1. Clones are available to other researchers.

The MAPMAKER program (Lander et al. 1987) was used to detect linkages and construct the map. Markers were assigned to likely groups using the GROUP command. Next, a first order approximation was made of the likely order; selected clones were then mapped using the three-point command and the other loci in the group added using the TRY command. A minimum LOD of 3.0 and a maximum recombination value (theta) of 0.25 was used. The Kosambi mapping function was used to determine recombination distances. We have now mapped 78 loci (Fig. 1) covering 291.6 cM using cDNA clones. RAPD markers and random genomic clones will be used to further cover the genome.

During mapping, we have made several interesting observations. First, segregation of codominant molecular markers in this population is expected to fit a 1:2:1 ratio. Of the 78 probes mapped to date, approximately 50% segregate normally ($p < 0.05$). The majority of the others have excess heterozygotes; for some clones, over 75% of the F₂ plants were heterozygous for the locus. The explanation of excess heterozygosity is not clear. Alfalfa suffers severely under inbreeding and must maintain heterozygosity in order to remain viable (Busbice and Wilsie 1966), and the loci which have excess heterozygosity may reflect the need for multiple alleles. Selection at any stage of development, from the zygote through the seedling, may have biased the progeny which we used as our population. The only stipulation we placed on our mapping plants was that they were sufficiently vigorous to produce enough tissue for DNA extraction. Less thrifty plants may have been homozygous at these loci. Most of the loci exhibiting segregation distortion were linked together (Groups 3, 8, 10, and 11).

Table 1. Molecular markers and nomenclature.

Institution	Type	Example name	Comments
Univ. of Georgia	cDNA	UGAc001	
Montana State Univ.	RAPD	MTS 1.1	relative marker size after decimal
Univ. of Wisconsin	RAPD genomic	MTST1.1 MTSg1; UWg1	telomere sequence multiple loci: UWg1a, UWg1b, etc.
Hungarian Academy of Sciences	cDNA	MTSc1; UWc1	
	genomic cDNA	fCG5 L23; U026; UGAc069	
	<i>Nodulin cDNAs:</i>		
		GSa2	glutamine synthetase
		LB12	leghemoglobin
		NMS25	nodulin-25
		NOD18B	nodulin-18
		NOD1G	nodulin-1G
		ENOD2	early nodulin-2
		NOD14	nodulin-14
		PEPC41	phosphoenol pyruvate carboxylase
	<i>Isozymes:</i>		
		<i>LAP-1, LAP-2</i>	leucine amino peptidase
		<i>PRX</i>	peroxidase
		<i>SKDH</i>	shikimic dehydrogenase
		<i>AAT2</i>	asparagine amino transferase
	<i>Morphological markers:</i>		
	Anthocyan	flower color	
	Dwarf	short growth	

These linkage groups probably represent real associations between markers because distortion at one locus will likely also be evident at other markers in the same chromosomal region.

Many cDNA clones are repeated in the genome. Of the total number of loci screened, 24% (80 of 322) are chloroplast or highly repeated nuclear clones. When probed on parental survey filters, they typically produce one or two monomorphic bands and occasionally several bands of lesser intensity. The latter bands may indicate secondary homology or sequences containing few repeat units. Of the remaining probes, 38% detect one or two bands, 26% detect three or four bands, and 36% detect more than four bands. Most clones in the latter two categories probably represent repeated sequences. The repeated clones are of limited value to mapping since the segregating loci in this mapping

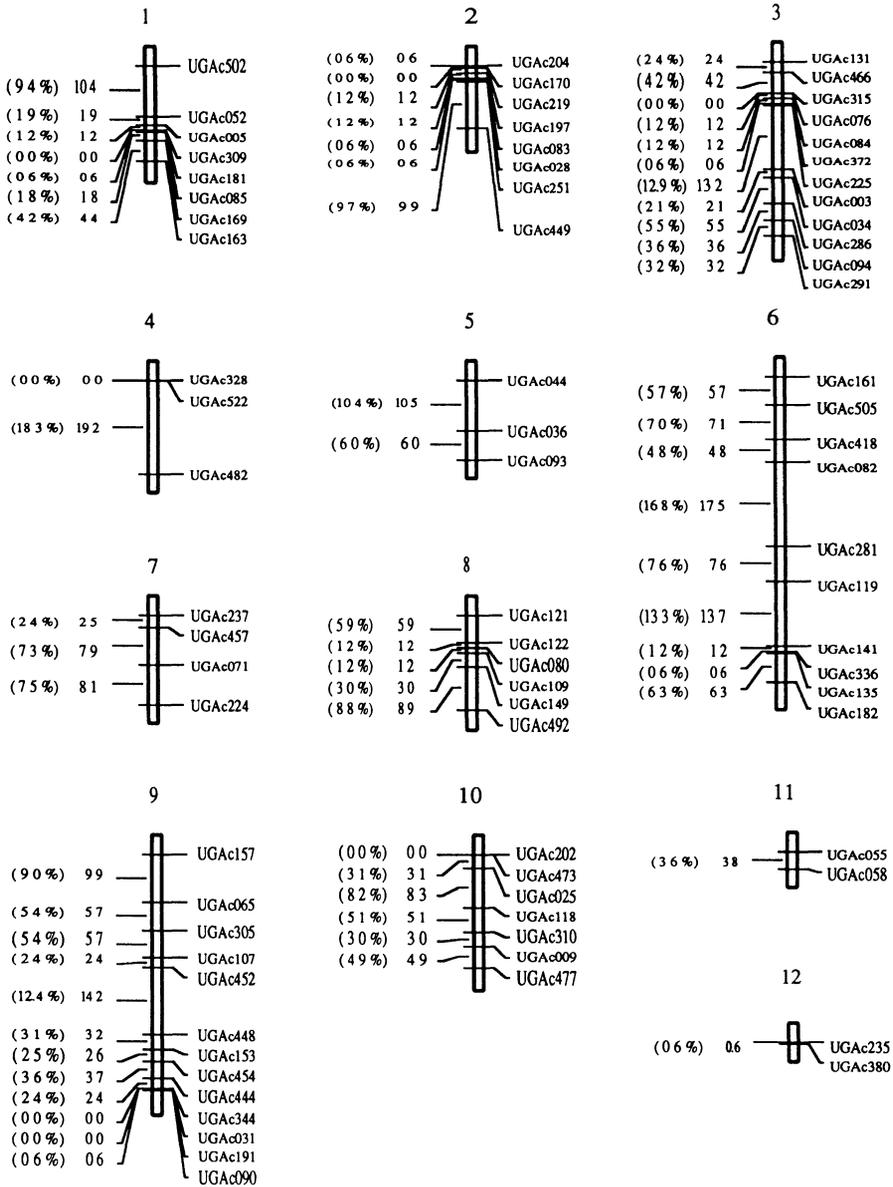


Fig 1 RFLP map of diploid alfalfa developed by the University of Georgia. Recombination fraction (%) and distance (cM) between markers are represented to the left of each chromosome, locus names are on the right

population may not be the same as those in another population. However, they may be useful for the production of skeleton linkage maps (Siracusa et al. 1991) or as tools to study genome duplication and architecture. A further problem is that some of the low-copy number clones don't segregate in a codominant fashion; some of these may be able to be mapped using a dominant marker (\pm)

system. For example, a band present in one parent, but not in the other, would be expected to segregate in a 3:1 manner in the F_2 because one could not distinguish between homozygotes for the band and heterozygotes (i.e. a band and a null). We are currently attempting these analyses.

Finally, although the cDNA clones have been chosen randomly, many of them cluster together, such as seen in linkage groups A and B. These clusters could be due to chance, to actual clustering of genes on alfalfa chromosomes, or to an artifact of this population such as limited recombination.

6. Mapping – Montana State University and University of Wisconsin

Our goal was to construct a linkage map that would be useful in cultivated germplasm. In order to obtain a segregating diploid population suitable for mapping, two CADL populations (Bingham and McCoy 1979) were selected on the basis of vigor and fertility and numerous pair-wise crosses were made between them. A single F_1 hybrid (2220-8) derived from one of these crosses was backcrossed to one of its parents (F_2 -16) to create the mapping population. This population was selected for mapping because of its general vigor, ploidy level stability, fertility, high germination rate, lack of visibly segregating lethal genes, and abundance of RFLP loci. The F_2 populations derived from the interpopulation crosses suffered from inbreeding depression and, therefore, did not meet these criteria and were not suitable as mapping populations.

DNA isolation using powdered, lyophilized leaf tissue was performed as reported (Echt et al. 1992; Kidwell and Osborn 1992). Restriction enzyme digestion, electrophoresis and Southern hybridization protocols were similar to those described for the University of Georgia with slight modifications. Random amplified polymorphic DNAs (RAPDs) were produced and scored as previously reported (Echt et al. 1992). Locus nomenclature is presented in Table 1.

A backcross involving allogamous (non-inbred) parents results in a maximum of three alleles at a locus. With both dominant and codominant DNA markers present, four single-locus segregation ratios can arise among the progeny: 1:1, 3:1, 1:2:1, and 1:1:1:1. (By contrast, an F_2 cross is limited to only two alleles per locus, resulting in only two single-locus DNA marker segregation ratios: 3:1 and 1:2:1.) Among the 1:1 segregants of an allogamous backcross, there are two different classes of loci, those which are homozygous in the P_1 parent and those which are homozygous in the P_2 parent. Recombination cannot be directly estimated between loci of the two different classes because the crossover and non-crossover genotypes of the progeny are indistinguishable. Paired loci from each 1:1 segregation class will always appear to be unlinked even though they both may show close linkage to a locus of one of the other segregation types.

Currently available genome mapping programs do not allow for analysis of loci of mixed segregation types of the 160 loci we analyzed, 107 were segregating 1:1, with 77 loci in one class and 30 in the other. We used the MAPMAKER

genome mapping program (Lander et al. 1987) to construct a map from the larger of the 1:1 segregant classes. The first-order function of MAPMAKER Macintosh v 1.0 was used ($r < 0.25$, $LOD > 4.0$) to determine linkage groups and approximate orders. The three-point mapping function was then used ($r < 0.3$, $LOD > 3.0$) on overlapping groups of 6–8 loci to confirm locus orders.

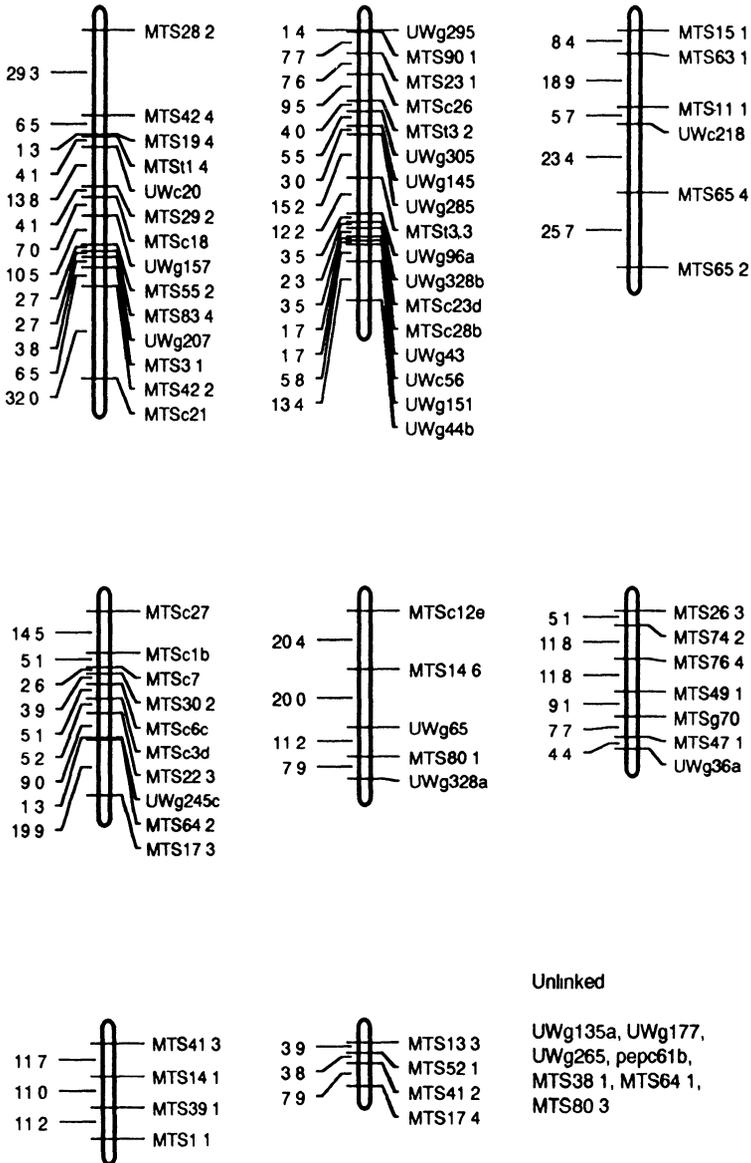


Fig 2 RFLP and RAPD map of diploid alfalfa developed by Montana State University and University of Wisconsin. Distances between markers (cM) are represented to the left of each chromosome, locus names are to the right

Final map distances were set for segment breaks to occur above a recombination fraction of 30%.

This map (Fig. 2) included 32 RFLP loci and 45 RAPD loci which mapped to eight linkage groups covering 628 cM using the Kosambi mapping function. Seven loci were unlinked. Similar to the University of Georgia data, 48% of the markers on this map exhibited significant ($p < 0.05$) segregation distortion. Loci having segregation distortion were not randomly distributed but tended to map together. The two largest linkage groups, having 14 and 16 loci respectively, had only one locus among them showing segregation distortion. The remaining six linkage groups were composed almost entirely of the remaining loci exhibiting segregation distortion. The mapping information and RFLP probes we have obtained are freely available to interested researchers.

7. Mapping – Hungarian Academy of Sciences, Szeged

A yellow flowered *M. sativa* subsp. *quasifalcata* plant (K93) was crossed to a purple flowered *M. sativa* subsp. *coerulea* plant (W2) to produce an F₁ hybrid. The F₁ was selfed and 138 F₂ individuals were used as the segregating mapping population. The parents are highly heterozygous for RFLP, isozyme, and morphological markers. Morphological mutants for leaf morphology, impaired nitrogen fixation (FIX), dwarfism, flower color, and seed storage proteins have been included in the mapping program. The experimental protocols are similar to those described above. The map was developed using the MAPMAKER program (Lander et al. 1987). Recombination fractions were determined using the Haldane mapping function. The current map consists of 39 markers (Table 1) and covers 232.9 cM (Fig. 3). Clones will be available to other researchers following publication of the map.

8. Application of molecular markers to alfalfa breeding and cultivar development

Most cultivar development programs proceed by initially selecting desirable plants to be used as parents. The selected plants are usually intermated in replicated crossing blocks (polycross) and the seed bulked. Several generations are then grown to produce enough seed for commercial release. Cultivars developed in this manner are termed synthetics (Hill and Elgin 1981). The number of parents which should be composited to create the new cultivar has been the subject of many studies (Hill and Elgin 1981; Busbice 1969; Hill 1971). Basically, with few parents, inbreeding can occur in the generations of seed increase so that the genetic composition of the released material is not the same as the experimentally tested germplasm. Conversely, the effect of inbreeding is diminished by selecting many plants as parents, but in doing so the combining

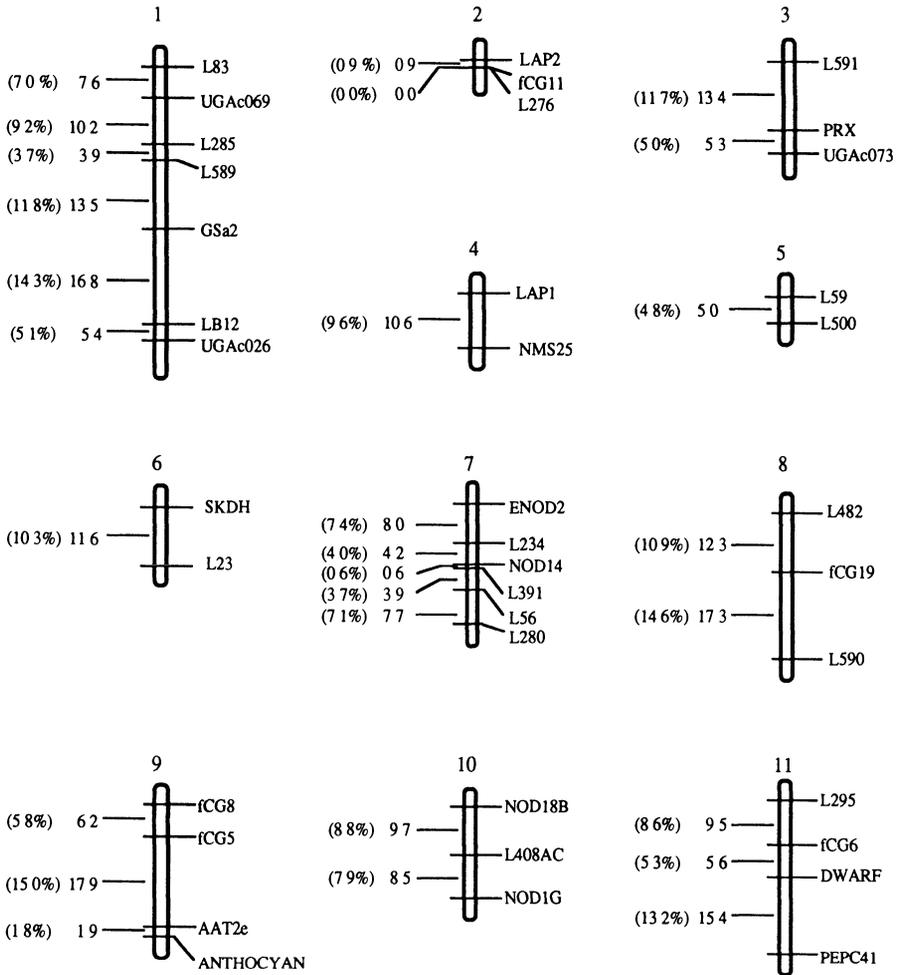


Fig 3 RFLP, isozyme, and morphological marker map of diploid alfalfa developed by the Hungarian Academy of Sciences. Recombination fraction (%) and distance (cM) between markers are represented to the left of each chromosome, locus names are on the right

ability of the plants with each other decreases (Busbice et al. 1972). In the latter case, any genetic variance not attributed to additive effects will not be utilized. As a result, Hill (1971) recommended that the selection of parents is more important than the combining ability of the selected parents if the number of parents is large.

Molecular marker-assisted selection is a potential aid to the selection process. After a saturated linkage map has been constructed and markers have been linked to traits of interest, indirect selection of traits based on the marker genotype is possible. Marker assisted selection allows a direct assay of the genotype, rather than simply the phenotype, of the selected plants. If the gene(s) controlling a trait are tagged with closely-linked markers, the heritability of the trait approaches one.

The gene action of autotetraploids has been actively researched, but the studies are necessarily indirect since until now no direct measure of the genotype was possible. Demarly (cited in Busbice et al. 1972) originally proposed the notion of either chromosome linkage groups or distinct loci enabling multiple alleles. Four alleles per locus are possible, a condition termed maximum heterozygosity (Dunbier and Bingham 1975). They developed populations with theoretically different proportions of multiallelic plants but the same gene frequencies. The populations with more multiallelic plants outyielded and had higher fertility and seed weight than the other populations. They concluded that maximum heterozygosity is important to realize the total potential of alfalfa. Using molecular markers, plants with exact genotypes can be selected so that the theory can be confirmed.

Commercial cultivar development methods do not enable production of maximally heterozygous cultivars without resorting to hybrid seed production, a route which has not proven feasible (Viands et al. 1988). Genetic studies using molecular markers can help resolve several problems. They can unequivocally prove whether maximum heterozygosity (that is, overdominance) is more important than additivity or dominance. Further, they can be used to identify which loci respond to maximum heterozygosity and which to other allele combinations. Knowing this, breeders can strive to select parents which will combine to produce the optimal allelic structure in the progeny, realizing that progeny will consist of an array of genotypes, at least after the generations of seed increase. The goal of the breeding program will be to screen parents so that the proportions of offspring which are grown in the farmers' fields will be skewed in a favorable manner.

Any breeding objective requires heritable variation for the trait of interest. Molecular markers and linkage maps can only assist in the selection of desirable plants; they do not enable development of cultivars with novel traits. Nevertheless, molecular markers are an extremely powerful tool to aid in the selection of traits traditionally difficult to select and to efficiently select many traits simultaneously. Marker-assisted selection relies on the ability to develop a population segregating for the trait of interest and to link the trait to molecular markers. Variation does not appear to be a problem in *M. sativa* (Brummer et al. 1991), so most populations which are segregating for a specific trait will also show segregation at most marker loci. Thus, no foreseeable problem exists in linking traits to markers. Once mapped, these traits may be selected indirectly using the linked markers.

Most practical breeding programs will use a molecular map primarily as a means to indirectly select genes tagged with molecular markers. Most currently available alfalfa cultivars have an array of resistances to various diseases, insects, and nematodes. Many of these resistances are conditioned by one or two gene systems; tagging of these genes with molecular markers should be relatively straightforward. Breeders can then simultaneously select multiple pest resistance genes, circumventing the costly and time-consuming conventional screening methods. Marker-assisted selection should facilitate the development

of cultivars highly resistant to many pests. Additionally, some pest resistances are not currently found within cultivated alfalfa. For example, no currently available cultivar has high resistance to the alfalfa weevil (*Hypera postica* Gyllenhal). Resistance has been identified in some annual species (*M. scutellata* (L.) Mill., *M. rugosa* Desr., *M. minima* Bart., and *M. disciformis*) as well as perennial species (*M. prostrata* Jacq. and *M. glandulosa* David.) (Sorensen et al. 1988). Field resistance to the alfalfa weevil has also been identified in perennial glandular-haired *Medicago* strains (Danielson et al. 1990). Molecular markers may be useful in identifying glandular hair genes (or other genes) which can be introgressed into cultivated alfalfa to provide increased resistance.

Other traits, notably yield, acid-soil tolerance, and quality, are not controlled by single genes, but rather are polygenic and influenced by the environment. Quantitative traits, such as these, are often difficult to select because the genotype is unknown and selection is based on the phenotype. Molecular markers offer the possibility of selection based on the genotype by identifying individual genes (quantitative trait loci or QTLs) which control the trait (Paterson et al. 1991). Identification of genes for quantitative traits, particularly when localized to particular environments and genetic backgrounds, will enable development of new cultivars tailored for certain conditions.

Alfalfa has little tolerance of poor soils, particularly acid soils (Devine et al. 1990), which are prevalent throughout the developing world. With an increased awareness of low input agriculture in the developed world, the ability of alfalfa to grow profitably in less than perfect conditions will be an important objective for breeders in the future. Quantitative loci for tolerance to low-phosphorus stress have been identified in maize (Reiter et al. 1991). Analogous procedures can be used in alfalfa for development of varieties tolerant to a variety of soil conditions. For example, by screening the USDA plant introduction collection, Bouton (unpub. data) has found germplasm which is more tolerant to aluminum and acid-soils than commercially grown cultivars. Mapping of genes controlling these traits would enhance the ability to integrate this low heritability trait into cultivars.

As breeders look outside the *M. sativa* complex for useful traits, the problem of introgressing undesirable genes along with desirable ones emerges. Fortunately, molecular markers provide a means to avoid linkage drag by selecting for cultivated germplasm marker alleles at all points on the map except for markers flanking the gene of interest. Young and Tanksley (1989) demonstrated that two generations of RFLP analysis would produce a fragment of the donor parent as small as that produced by 100 generations of traditional backcrossing. With such a savings of time and money, breeders cannot overlook the importance of genetic mapping for introgression studies.

Another, long-term application of the alfalfa genome map will be map-based cloning of genes. Both basic genetic studies and practical breeding objectives may utilize gene cloning. One of the most interesting features of alfalfa, and other legumes, is their symbiosis with nitrogen fixing bacteria. The plant genes involved in this symbiosis are crucial to the understanding of this subject. One

possible method to clone some of these genes is through chromosome walking or jumping from linked markers (Poustka et al. 1986). Kiss and his colleagues are planning to use alfalfa maps to locate and clone several genes he has identified. Also, cloning of resistance genes from annual species which do not cross with cultivated alfalfa may be a prerequisite for the introduction of these genes into cultivated germplasm by transformation.

9. Problems of molecular marker analysis in alfalfa

Three research groups are currently producing molecular maps of alfalfa using diploid populations. Virtually all useful applications of the maps to practical breeding programs are predicated on the ability to transfer the diploid map to segregating tetraploid populations. Although theoretically possible, actually mapping with four possible alleles rather than two may be confusing. More progeny will need to be phenotyped and genotyped in segregating tetraploid populations than necessary at the diploid level. The maximum number of possible genotypes at a single locus in a population developed by selfing a maximally heterozygous plant is three in diploids and thirty-five in tetraploids with double reduction or nineteen in tetraploids without double reduction. Additionally, distinguishing among many genotypes will be difficult when only allele copy numbers differ (e.g. $A_1 A_1 A_1 A_2$ and $A_1 A_1 A_2 A_2$, where A_1 and A_2 are two hypothetical alleles at a locus. Wu et al. (1992) recently reported a method of estimating linkage in polyploids using single-dose restriction fragments. This method may need to be adopted by alfalfa geneticists for mapping at the tetraploid level.

Another problem with mapping in alfalfa is the extreme heterozygosity among plants. As more parents are chosen to use in a cultivar development program, the likelihood of a new allele at the marker locus of interest increases. Linkage of the trait to the new allele is not known. Thus, marker-assisted selection may be difficult if multiple alleles are present in the population and the linkage of the alleles to the trait is not fully understood. The use of RAPD (Williams et al. 1990) or RFLP markers which detect multiple loci may have limited value since the mapped bands in one population may not be the same bands which segregate or are mapped in another population.

10. Acknowledgements

Georgia: Thanks to Ann Bunce for excellent greenhouse assistance. Research supported by a grant from the University of Georgia Research Foundation.

Montana, Wisconsin: Thanks to Linda Erdahl (MSU) for technical help. This work was supported at MSU by a grant from Pioneer Hi-Bred International and by USDA grant #9101510, and at UW by a Pioneer Graduate Fellowship (to K.K.) and the College of Agricultural and Life Sciences

Hungary: Thanks to Mrs. Gyöngyi Somkuti for skillful technical assistance.

11. References

- Aarts, E and Korst, J (1989) Simulated Annealing and Boltzmann Machines John Wiley, New York
- Apuya, N R , Frazier, B L , Keim, P , Roth, E J and Lark, K G (1988) Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* (L) merrill Theor Appl Genet 75 889-901
- Barnes, D K , Golpen, B P and Baylor, J E (1988) Highlights in the USA and Canada In A A Hanson, D K Barnes and R R Hill (eds), Alfalfa and Alfalfa Improvement, pp 1-24 ASA, Madison, WI
- Bingham, E T (1971) Haploids from cultivated alfalfa, *Medicago sativa* L Nature 221 865-866
- Bingham, E T (1991) Registration of isogenic populations of diploid and tetraploid alfalfa, W2xiso-1 and W4xiso-1 Crop Sci 31 496
- Bingham, E T and McCoy, T J (1979) Cultivated alfalfa at the diploid level origin, reproductive stability, and yield of seed and forage Crop Sci 19 97-100
- Bolton, J L (1962) Alfalfa Botany, Cultivation, and Utilization Interscience Publishers, Inc , New York
- Brunner, E C , Kochert, G and Bouton, J H (1991) RFLP variation in diploid and tetraploid alfalfa Theor Appl Genet 83 89-96
- Busbice, T H (1969) Inbreeding in synthetic varieties Crop Sci 9 601-604
- Busbice, T H and Wilsie, C P (1966) Inbreeding depression and heterosis in autotetraploids with application to *Medicago sativa* L Euphytica 15 52-67
- Busbice, T H , Hill, R R Jr and Carnahan, H L (1972) Genetics and breeding procedures In C H Hanson (ed), Alfalfa Science and Technology, pp 263-318 ASA, Madison, WI
- Danielson, S D , Manglitz, G R and Sorenson, E L (1990) Field resistance of perennial glandular-haired *Medicago* strains and alfalfa cultivars to the alfalfa weevil Euphytica 49 177-182
- Dellaporta, S L , Wood, J and Hicks, J B (1983) A plant DNA miniprep preparation Version II Plant Mol Biol Rep 1 19-21
- Devine, T E , Bouton, J H and Mabrahtu, T (1990) Legume genetics and breeding for stress tolerance and nutrient efficiency In V C Baligar and R R Duncan (eds), Crops as Enhancers of Nutrient Use, pp 211-252 Academic Press, New York
- Dunbier, M W and Bingham, E T (1975) Maximum heterozygosity in alfalfa Results using haploid-derived autotetraploids Crop Sci 15 527-531
- Echt, C S , Erdahl, L A and McCoy, T J (1992) Genetic segregation of random amplified polymorphic DNA in diploid cultivated alfalfa Genome 35 84-87
- Hill, R R Jr (1971) Effect of the number of parents on the mean and variance of synthetic varieties Crop Sci 11 283-286
- Hill, R R Jr and Elgin, J H Jr (1981) Effect of the number of parents on performance of alfalfa synthetics Crop Sci 21 298-300
- Kidwell, K K and Osborn, T C (1992) Simple plant DNA isolation procedures In J Beckmann and T C Osborn (eds), Plant Genomes Methods for Genetic and Physical Mapping, pp 1-13 Kluwer Academic Publishers, Dordrecht, The Netherlands
- Lander, E S , Green, P , Abrahamson, J , Barlow, A , Daly, M J , Lincoln, S E and Newburg, L (1987) MAPMAKER an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations Genomics 1 174-181
- Lesins, K and Gillies, C B (1972) Taxonomy and Cytogenetics of *Medicago* In C H Hanson (ed) Alfalfa Science and Technology, pp 53-86 ASA, Madison, WI
- Lesins, K and Lesins, I (1979) Genus *Medicago* A Taxogenetic Study Dr W Junk, The Hague
- McCoy, T J and Bingham, E T (1988) Cytology and cytogenetics of alfalfa In A A Hanson, D K Barnes and R R Hill (eds), Alfalfa and Alfalfa Improvement, pp 737-776 ASA, Madison, WI
- McCoy, T J and Smith, L Y (1986) Interspecific hybridization of perennial *Medicago* species using ovule-embryo culture Theor Appl Genet 71-772-783
- Michaud, R , Lehman, W F and Rumbaugh, M D (1988) World distribution and historical

- development In A A Hanson, D K Barnes and R R Hill (eds), *Alfalfa and Alfalfa Improvement*, pp 25-91 ASA, Madison, WI
- Murray, M G and Thompson, W F (1980) Rapid isolation of high molecular weight plant DNA *Nucl Acids Res* 8 4321-4325
- Oste, C (1988) Polymerase chain reaction *Biotechniques* 6 162-167
- Paterson, A H, Damon, S, Hewitt, J D, Zamir, D, Rabinowitch, H D, Lincoln, S E, Lander, E S and Tanksley, S D (1991) Mendelian factors underlying quantitative traits in tomato Comparison across species, generations, and environments *Genetics* 127 181-197
- Poustka, A, Pohl, T, Barlow, D P, Zehetner, G, Craig, A, Michiels, F, Ehrlich, E, Frischauf, A - M and Lehrach, H (1986) *Molecular Approaches to Mammalian Genetics*, pp 131-139 Cold Spring Harbor Symposia on Quantitative Biology, Vol LI Cold Spring Harbor Laboratory
- Quiros, C F (1982) Tetrasomic segregation for multiple alleles in alfalfa *Genetics* 101 117-127
- Quiros, C F (1983) Alfalfa, luzerne (*Medicago sativa* L.) In S D Tanksley and T J Orton (eds), *Isozymes in Plant Genetics and Breeding, Part B*, pp 253-294 Elsevier, Amsterdam
- Quiros, C F and Baughan, G R (1988) The genus *Medicago* and the origin of the *Medicago sativa* complex In A A Hanson, D K Barnes and R R Hill (eds), *Alfalfa and Alfalfa Improvement*, pp 93-124 ASA, Madison, WI
- Reiter, R S, Coors, J G, Sussman, M R and Gabelman, W H (1991) Genetic analysis of tolerance to low-phosphorus stress in maize using restriction fragment length polymorphisms *Theor Appl Genet* 82 561-568
- Saghai-Marooif, M A, Soliman, K M, Jorgensen, R A and Allard, R W (1984) Ribosomal DNA spacer-length polymorphisms in barley Mendelian inheritance, chromosomal location, and population dynamics *Proc Natl Acad Sci USA* 81 8014-8018
- Sangduen, N, Sorensen, E L and Liang, G H (1982) A perennial \times annual *Medicago* cross *Can J Genet Cytol* 24 361-365
- Shure, M, Wessler, S and Federoff, N (1983) Molecular identification of the *Waxy* locus in maize *Cell* 35 225-233
- Siracusa, L D, Jenkins, N A and Copeland, N G (1991) Identification and applications of repetitive probes for gene mapping in the mouse *Genetics* 127 169-179
- Sorensen, E L, Byers, R A and Horber, E K (1988) Breeding for insect resistance In A A Hanson, D K Barnes and R R Hill (eds), *Alfalfa and Alfalfa Improvement*, pp 859-902 ASA, Madison, WI
- Stanford, E H (1951) Tetrasomic inheritance in alfalfa *Agron J* 43 222-225
- Vians, D R, Sun, P and Barnes, D K (1988) Pollination control Mechanical and sterility In A A Hanson, D K Barnes and R R Hill (eds), *Alfalfa and Alfalfa Improvement*, pp 931-960 ASA, Madison, WI
- Williams, J G K, Kubelk, A R, Livak, K J, Rafalski, J A and Tingey, S V (1991) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers *Nucl Acids Res* 18 6531-6535
- Wu, K K, Burnquist, W, Sorrells, M E, Tew, T L, Moore, P H and Tanksley, S D (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragments *Theor Appl Genet* 83 294-300
- Young, N D and Tanksley, S D (1989) Restriction fragment length polymorphism maps and the concept of graphical genotypes *Theor Appl Genet* 77 95 101

9. An integrated RFLP map of *Arabidopsis thaliana*

HOWARD M. GOODMAN¹, SUSAN HANLEY¹, SAM CARTINHOOR¹,
J. MICHAEL CHERRY¹, BRIAN HAUGE¹, ELLIOT MEYEROWITZ²,
LEONARD MEDRANO², SHERRY KEMPIN², PIET STAMM³ and
MAARTEN KOORNNEEF³

¹ Departments of Genetics, Harvard Medical School and Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, U.S.A.; ² Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, U.S.A.; ³ Department of Genetics, Agricultural University of Wageningen, Dreijenlaan 2, 6703 HA Wageningen, The Netherlands

Over the past several years, *Arabidopsis thaliana* has gained increasing popularity as a model system for the study of plant biology. Its short life cycle, small size and large seed output make it well suited for classical genetic analysis (reviewed in Meyerowitz 1987). Mutations have been described affecting a wide range of fundamental developmental and metabolic processes (reviewed in Estelle and Somerville 1986). A genetic linkage map consisting of some 90 loci has been assembled (Koornneef 1987) and an increasing number of cloned genes are available. In addition, *Arabidopsis* is ideally suited for physical mapping studies since it has a very small genome (approximately 100,000 kb; Hauge and Goodman, unpub. result) and a remarkably low content of interspersed repetitive DNA (Pruitt and Meyerowitz 1986). The availability of a complete physical map of the *Arabidopsis* genome will greatly simplify the cloning of any gene based solely on its mutant phenotype and genetic map location. For the map to be of any utility it is necessary to align the physical map with the classical genetic linkage map via an RFLP map.

The current RFLP map of *Arabidopsis thaliana* is presented in Table 1. Using methods essentially equivalent to the two previously published RFLP maps (Chang et al. 1988; Nam et al. 1989) the number of markers in each data set have been increased and the independent data sets have been mathematically integrated into a single map using the computer program 'Joinmap' (Stam 1993). It should be noted, however, that both the resolution of the RFLP map and the alignment of the RFLPs with the genetic linkage map are somewhat limited. Therefore, it is currently difficult to assign with confidence a precise position on the classical genetic map for a given RFLP probe since only a few genetic markers per chromosome were used to align the RFLP map with the genetic linkage map. In addition, due to the necessarily limited number of progeny used for the segregation analysis, we estimate that the resolution of the RFLP map is only on the order of 2 cM. In our early mapping efforts we primarily worked with randomly selected bacteriophage λ (Chang et al. 1988) or cosmid clones (Nam et al. 1989). Subsequent efforts have mainly been devoted to mapping cloned genes obtained from various laboratories as indicated in Table 1. The majority of genes that have been mapped have not been genetically

identified. In other words their cognate mutation is not known. In a few cases both the gene and the genetic locus is known. For example, *ag* (agamous) (Yanofsky et al. 1990), *ap3* (apetala 3) (Jack et al. 1992), and *tt-3* and *tt-5* (transparent testa 3 & 5) (Shirley et al. 1992) have been cloned. The mapping of these loci provide immediate contact points between the physical map and the classical genetic linkage map. As new loci are mapped with respect to the RFLPs, additional contact points will be established and the map will be refined.

Table 1 Classical genetic markers are indicated in italics (Koorneef, 1987) and classical genetic markers which have also been cloned and RFLP mapped are indicated in bold type. These are as follows: *chr1 an*, *angustifolia*, **chl3**, chlorate resistant, *chl*, chlorina, *apl*, *apetala*, *gl2*, *glabra*, *clv1*, *clavata*, **ga2**, gibberellin requiring, *chr2 hy3*, long hypocotyl, *cp2*, compacta, *er*, erecta, *as*, asymmetric leaves and lobed leaves, *py*, pyrimidine requiring, *cer8*, eceriferum, *chr3 hy2*, long hypocotyl, *dwf*, auxin resistant dwarf, **abi3**, abscisic acid insensitive, *gl1*, *glabra*, **ap3**, *apetala*, **tt5**, transparent testa, *chr4 ga1*, gibberellin requiring, *bp*, brevipedicellus, **ag**, *agamous*, *cer2*, eceriferum, *ap2*, *apetala*, and *chr5 tt4*, transparent testa, *pi*, *pistillata*, *ttg*, transparent testa *glabra*, **tt3**, transparent testa, *tz*, thiozole requiring, *biol1*, biotin auxotroph. All other markers in plain text have been cloned and RFLP mapped, but not correlated with a mapped mutation. RFLP markers with names of the form m-# are random bacteriophage λ clones from the Meyerowitz laboratory (Chang et al. 1988) previously designated either λ At-# or LEM-#, names of the form pCITd-#, pCIT-# or pCITN7-- are random cosmid clones from the Meyerowitz laboratory, and names of the form g-# are random cosmid clones from the Goodman laboratory (Nam et al. 1989). The sources of the other markers can either be found in the publically available database AAtDB (Cherry et al. 1993) or will be published elsewhere (Hauge et al. 1993). The numbers under the cM heading indicate the position on the chromosome using the Kosambi mapping function in Joinmap.

CHR1	cM	CHR2	cM	CHR3	cM	CHR4	cM	CHR5	cM
pvv4	0	pCITd112a	0	myb	0	g6844	0	pCITd94	0
g21491	6	m246	10.8	pCITd39	2.6	g3843	4	g21488	1.2
<i>an</i>	14.5	pCIT1291	13.2	GAP-C	2.7	U2 9snRNA	6	ubq6121	2.3
m488	23.7	g4133	14.7	m302	3.5	gal-14	7.5	pAt80	3.9
g5972	25.1	g4532	17	CD06119	4.2	gal	11.3	m562	5.4
pCITd91	25.4	m497	17.7	m262	5.4	<i>bp</i>	13.5	m447	7.3
IPhAra1	26.2	g4553	18.9	m472	7.6	auxinBP	15.6	g3715	8.2
GTPbp	27.7	a14G4	31.8	g3838	8.9	g2616	16.2	CD05629	10.4
g4715	28	pCITd100	34.6	m583	9.4	m506	17.5	m217	10.9
1a8	28.1	hy3	34.9	peaf	10.5	m456	18.8	KG-31	13.3
g19821	28.6	m216	36.4	g4119	11.7	m448	22.2	U2 5snRNA	13.4
m322	28.8	m104	39.2	g17341	12.5	PK-87	23.8	g3837	14.2
g19857	29.7	PK-20	39.4	m243	12.6	pCITf3	28.6	ASA-1	14.5
g5957	29.7	m605	40.6	g4523	13.3	pCITd23	29.1	PBS811	16.3
m333	29.8	<i>cp2</i>	42.4	hsp70-9	14.7	m518	31.4	m224	17.4
m241	31.7	m465	44.4	<i>hy2</i>	14.8	m210	37.4	g6830	18
m219	33.3	m251	46.7	g4547	16.1	g6837	38.8	pCIT1243	19.9
g21497	35.1	Gpa1	55.4	g2488-a	16.2	m326	40.1	g3021	20
g2358	38.6	g6842	56.2	pCITf7P	18.3	m580	41.5	tt4	20.1
pCITN7-31	39.2	<i>er</i>	57.1	m228	19.1	g10086	42.1	pCITd37	21.3
g3786	39.4	PBS707	60.8	m317	19.3	g4565-a	43.8	g5962	22.3
m235	44.6	m220	62.2	KG-17	19.8	m226	44.3	g4568	23.8
g2395	45.8	m283	63.3	m560	22.4	pCITd71	45.9	pCIT718	25.5
g3829	47.3	ASA-2	63.6	<i>dwf</i>	22.5	ag	48.7	AR119	25.8
g17286	50.7	g21502	65.7	GS-KB6	22.6	g4539	49.9	pCITf16	27
m215	51.2	m323	65.8	g4708	23.1	g3845	50.9	pGATC-11	27
m310	51.5	m429	68.7	pCIT1240	24.1	g19833	53	g4111	27.2
g19834	53.9	g6191	69.2	g6220	26.3	g19838	53.2	g4560	27.5

References

- Chang, C., Bowman, J.L., DeJohn, A.W., Lander, E.S. and Meyerowitz, E.M. (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. U.S.A. 85: 6856–6860.
- Cherry, J.M., Cartinhour, S. and Goodman, H.M. (1992) AAtDB, an *Arabidopsis thaliana* database. Plant Mol. Biol. Rep. 10: 308–309, 409–410.
- Estelle, M.A. and Somerville, C.R. (1986) The mutants of *Arabidopsis*. Trends Genet. 2: 89–93.
- Hauge, B.M., Hanley, S.M., Cartinhour, S., Cherry, J.M., Goodman, H.M., Koornneef, M., Stam, P., Chang, C., Kempin, S., Medrano, L. and Meyerowitz, E.M. (1993) An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome. Plant J. 3: 745–754.
- Jack, T., Brockman, L.L. and Meyerowitz, E.M. (1992) The homeotic gene APETALA3 of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. Cell 68: 683–697.
- Koornneef, M. (1987) Linkage map of *Arabidopsis thaliana* (2n = 10). In: S.J. O'Brien (ed.), Genetic Maps, pp. 742–745. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Meyerowitz, E.M. (1987) *Arabidopsis thaliana*. Ann. Rev. Genet. 21: 93–111.
- Nam, H.-G., Giraudat, J., den Boer, B., Moonan, F., Loos, W.D.B., Hauge, B.M. and Goodman, H.M. (1989) Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. Plant Cell 1: 699–705.
- Pruitt, R.E. and Meyerowitz, E.M. (1986) Characterization of the genome of *Arabidopsis thaliana*. J. Mol. Biol. 187: 169–183.
- Shirley, B., Hanley, S. and Goodman, H.M. (1992) Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis transparent testa* mutations. Plant Cell 4: 333–347.
- Stam, P. (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. Plant J. 3: 739–744.
- Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A. and Meyerowitz, E.M. (1990) The protein encoded by *Arabidopsis* homeotic gene *agamous* resembles transcription factors. Nature 346: 35–39.

10. RFLP maps of barley

ANDRIS KLEINHOF¹ and ANDRZEJ KILIAN^{1,2}

¹ Departments of Crop & Soil Sciences and Genetics & Cell Biology, Washington State University, Pullman, WA 99164-6420, U S A , ² Department of Genetics, Silesian University, 40-032 Katowice, Poland

Contents

1 Introduction	163	4 4 Chromosome 4 (4H)	180
2 Crosses and doubled haploid populations	164	4 5 Chromosome 5 (1H)	183
3 RFLP markers	165	4 6 Chromosome 6 (6H)	185
4 RFLP maps	167	4 7 Chromosome 7 (5H)	189
4 1 Chromosome 1 (7H)	168	5 Anomalous probes	192
4 2, Chromosome 2 (2H)	173	6 Summary	194
4 3 Chromosome 3 (3H)	176	7 Acknowledgements	194
		8 References	195

1. Introduction

Barley, *Hordeum vulgare* L., is among the oldest cultivated crops dating 5000 to 7000 years B.C. (Clark 1967; Harlan 1976, 1979) and perhaps much older (Wendorf et al. 1979). Barley has also been a favorite genetic experimental organism since the rediscovery of Mendel's laws of heredity (Tschermak 1901; cited from Smith 1951). The widespread use of barley is attributable to its diploid nature ($2n = 2x = 14$), self fertility, large chromosomes (6–8 μm), high degree of natural and easily inducible variation, ease of hybridization, wide adaptability, and relatively limited space requirements. The early literature on the genetics of barley and related information has been reviewed (Smith 1951; Nilan 1964, 1974; Briggs 1978) and more recently in Rasmusson (1985) and Shewry (1992). A comprehensive listing and description of the genes known at the time can be found in Sogaard and von Wettstein-Knowles (1987), updated by Von Wettstein-Knowles (1992). Periodic publications such as the Proceedings of International Barley Genetic Symposia (Lamberts et al. 1964; Nilan 1971; Gaul 1976; Asher et al. 1981; Yasuda and Konishi 1987; Munck 1992) and the annual Barley Genetics Newsletter (since 1971) summarize and update the barley genetic literature.

The application of molecular techniques to barley genetic mapping started

slowly, perhaps due to its large genome size, $1C = 5.3 \times 10^9$ bp (Bennett and Smith 1976). Several advantages, however, make barley the organism of choice among the large genome types for molecular mapping. These include the ability to develop doubled haploid lines by several methods and the availability of cytogenetic stocks, particularly the barley-wheat addition lines.

Doubled haploid lines can be developed by two techniques which sample either female or male gametes. One method is based on the observations that when *Hordeum vulgare* female flowers are pollinated with *Hordeum bulbosum*, initial fertilization and zygote formation takes place, but the *H. bulbosum* chromosomes are subsequently eliminated in the hybrid embryos (Kasha and Kao 1970; Kasha 1974). The haploid embryos can be rescued and homozygous plants developed by tissue culture and chromosome doubling techniques (Chen and Hayes 1989; Kasha 1974). The resulting doubled haploid lines are random samples of female gametes. The North American Barley Genome Mapping Project (NABGMP) has used this technique to prepare doubled haploid lines from three crosses (described later). Alternatively, anther culture can be used to develop doubled haploid lines that represent random samples of male gametes. This technique has been used to develop several doubled haploid populations for mapping of the barley genome (Heun et al. 1991; Graner et al. 1991).

The availability of barley-wheat disomic (Islam 1980) and ditelosomic (Islam 1983) addition lines further facilitates barley genetic studies. These lines contain a normal complement of Chinese Spring wheat chromosomes and a cv. Betzes barley chromosome or chromosome arm. All individual barley chromosome addition lines, except chromosome 5, are fertile and reasonably stable. These lines have been used to organize individual markers and linkage groups into chromosomes and identify centromere locations (Heun et al. 1991; Graner et al. 1991; Kleinhofs 1992; this report).

The first barley RFLP map published was for chromosome 6 (Kleinhofs et al. 1988) followed by a partial map of the whole genome incorporating RFLP, morphological, isozyme and polymerase chain reaction generated markers (Shin et al. 1990). Recently, three more extensive maps of the barley genome have been generated (Heun et al. 1991; Graner et al. 1991; Kleinhofs 1992). These maps and other less extensive efforts, jointly represent approximately 600 RFLP markers mapped to the barley genome. To date, only a few common markers have been placed on the different maps. Unfortunately, this means that it is almost impossible to merge these maps into one extensive map of the barley genome at this time.

2. Crosses and doubled haploid populations

The current barley RFLP maps are based on four segregating populations. These are Steptoe by Morex, 150 doubled haploid lines developed by the North American Barley genome Mapping Project (NABGMP) using the *H. bulbosum* technique (Hayes, unpub.); Proctor by Nudinka, 113 doubled haploid lines

developed by the anther culture technique (Heun et al. 1991); Igri by Franka, 71 doubled haploid lines developed by the anther culture technique (Graner et al. 1991); and Vada by *H. vulgare* subsp. *spontaneum* line 1B-87, consisting of 135 F₂ individuals (Graner et al. 1991). Other materials used to develop partial maps are a population of 58 F₂ individuals from a cross of the Steptoe nitrate reductase-deficient mutant Az12 by Winer and a population of 147 F₂ individuals from a cross *H. vulgare* subsp. *vulgare* cv. Aramir by *H. vulgare* subsp. *spontaneum*. The Steptoe mutant Az12 by Winer derived population was used to map chromosomes 1 and 6 (Kleinhofs et al. 1988; Kilian et al., unpub.) and the Aramir by *H. spontaneum* derived population was used to map chromosome 4 (Hinze et al. 1991).

Maps discussed in this paper are sometimes designated and referred to, for brevity, by abbreviations of the parental cultivars crossed to produce the population used for mapping i.e. S/M – Steptoe by Morex; P/N – Proctor by Nudinka; I/F – Igri by Franka; V/H.s. – Vada by *H. spontaneum* line 1B-87; S/W – Steptoe mutant Az12 by Winer.

The NABGMP has developed two additional doubled haploid populations consisting of at least 150 lines each that are available for mapping. These populations were produced using the *H. bulbosum* method from the crosses Harrington by TR306 and Morex by Harrington. The parents for the NABGMP populations were selected based on preliminary polymorphism analyses and their agronomic characteristics. Cvs. Morex (Rasmusson and Wilcoxson 1979) and Harrington are the North American brewing industries 6-rowed and 2-rowed malt standards, respectively. Cvs. Steptoe (Muir and Nilan 1973) and TR306 are high yielding 6-rowed and 2-rowed feed barleys, respectively. The Steptoe by Morex and Morex by Harrington parents and doubled haploid lines are available from P. Hayes (Oregon State Univ., Corvallis, OR, U.S.A.) and the Harrington by TR306 parents and doubled haploid lines are available from K. Kasha (Univ. Guelph, Guelph, Ontario, Canada).

3. RFLP markers

Probes from many different sources have been used to map the barley genome. These are listed in Tables 1–7 and summarized in Table 8. The RFLP markers are either anonymous cDNA or genomic DNA probes or cloned known-function probes (mostly cDNA). Probes from species other than barley such as wheat, *Triticum tauschii*, oat and some maize sequences hybridize well with barley DNA and have also been used. The known-function gene probes are designated consistent with previously used or suggested gene symbols or, if none, a symbol is assigned following established guidelines (Barley Genet. Newsletter 11: 1–16, 1981). Donors of these probes and appropriate references are identified (Tables 1–7). In cases where it was not possible to determine which previously used locus designation corresponded to the locus uncovered by the

RFLP probe, a new locus number was used. For example, we mapped five alcohol dehydrogenase loci using an alcohol dehydrogenase cDNA probe (Good et al. 1988), but could not resolve which of the RFLP loci corresponded to *Adh* 1, 2 and 3 loci identified by isozyme analyses (Sogaard and von Wettstein-Knowles 1987). Thus, we designated the RFLP loci *Adh* 4, 5, 6, 7, and 8. Similar problems arose with the peroxidase loci. Here the gene symbol *Prx* is used for the RFLP identified loci to differentiate them from the more commonly used isozyme symbol *Per*. These loci designations will have to be reevaluated as new information is accumulated.

Although rules for naming loci detected with anonymous probes have not yet been established for barley, the same rules as used for naming genes should be followed where appropriate. A three letter (capital) designation followed by an arbitrary number is commonly used. For example ABC156 and ABG 472 stand for American Barley cDNA No. 156 and American Barley genomic DNA No. 472, respectively. If a probe detects several loci, they are designated with a capital letter following the arbitrary number. For example, several polymorphic bands detected by the probe ABC156 would be designated ABC156A, ABC156B, and so forth. A small letter has been used as a prefix to the gene symbol to identify loci detected as isozymes (i) or morphological markers (m) by the NABGMP. In one case the prefix 'c' was used to designate cDNA probes (Graner et al. 1991).

The source of the anonymous cDNA probes and their designations are as follows. The NABGMP group anonymous cDNA probes are from cv. Steptoe leaf or Steptoe malted seed libraries produced by C. Huang (in A. Kleinhofs' laboratory) and R. Skadsen (Cereal Crops Research Unit, Madison, WI, U.S.A.), respectively. These probes are designated ABC (American Barley cDNA) followed by an arbitrary number.

The Cornell probes are designated as in Heun et al. (1991); BCD for barley, cv. Willis, leaf cDNA and CDO for oat, cv. Brooks, leaf cDNA followed by an arbitrary number. The NABGMP group has used several of the Cornell probes for mapping the S/M cross. In those cases where the location of the marker appeared to be different in the S/M map from the P/N map, the probe designation was altered by adding a capital letter after the arbitrary number to suggest a different locus may be involved.

The anonymous probes used to map the I/F and V/H.s. crosses are designated as in Graner et al. (1991); cMWG for barley leaf cDNA (cv. not identified). The lower case 'c' indicates cDNA while the letters 'MWG' refer to the authors' institutes. The MWG group used a lower case letter to designate multiple loci detected by a single probe. We recommend that these be changed to capital letter to correspond with the designations used by Heun et al. (1991) and the NABGMP group and to avoid confusion with the use of the lower case letter to designate alleles in barley genetic symbols (Barley Genet. Newsletter 11: 1-16, 1981).

The Institute of Plant Science Research anonymous probes are from wheat cv. Chinese Spring leaf cDNA library. These are designated PSR followed by an arbitrary number (Chao et al. 1988; Sharp et al. 1988).

All of the anonymous genomic DNA probes described herein are from *Pst*I libraries. The library used by the NABGMP group was prepared by N. Lapitan (Dept. Agronomy, Colorado State Univ., Fort Collins, CO, U.S.A.). These markers are designated ABG (American Barley Genomic DNA) followed by an arbitrary number. Cornell University wheat genomic DNA probes are designated WG followed by an arbitrary number (Heun et al. 1991). Some barley genomic DNA markers used by Heun et al. (1991) are designated BG. Other barley genomic probes, used by Graner et al. (1991) to map the I/F and V/H.s. crosses are designated MWG followed by an arbitrary number. *Triticum tauschii* genomic DNA probes, provided by B.S. Gill (Gill et al. 1991), have been used to map the P/N and S/M crosses (Heun et al. 1991; Kleinhofs 1992). These are designated ksu (Kansas State University) followed by a capital letter and an arbitrary number. Heun et al. (1991), also include the letters DG immediately after 'ksu' and before the letter designation to identify these clones as arising from the wheat D genome.

4. RFLP maps

Cultivated barley (*Hordeum vulgare* subsp. *vulgare*) has seven pairs of distinct chromosomes designated by Arabic numbers 1–7 (Nilan 1964; Ramage 1985). The five chromosomes without satellites are designated 1 to 5 based on their relative lengths measured at mitotic metaphase, with chromosome 1 being the longest and chromosome 5 the shortest. The two chromosomes with satellites are designated 6 and 7, with chromosome 6 having the larger satellite and being shorter than chromosome 7 which has the smaller satellite. The initial association of linkage groups with specific chromosomes was accomplished using translocations (Burnham and Hagberg 1956) and trisomics (Tsuchiya 1960). The ranking of the lengths of the three longest chromosomes has been reported to be different from the original assignments in some studies; however, to avoid confusion, the originally used numbers have been retained by barley geneticists. Each of the seven barley chromosomes can be identified by its distinctive Giemsa C- and N-banding pattern (reviewed in Ramage 1985; Linde-Laursen and Jensen 1992). Based on the few reports of barley pachytene chromosome analyses (Sarvella et al. 1958; Singh and Tsuchiya 1975), the relative lengths and arm ratios agree reasonably well with those published for mitotic metaphase chromosomes.

Several barley chromosomes are metacentric, particularly chromosome 1, resulting in some confusion of the 'short' and 'long' arm designations (Singh and Tsuchiya 1982). Islam (1983) avoided this problem by designating chromosome 1 ditelosomic addition lines 1α and 1β . We favor the proposal of Linde-Laursen and Jensen (1992) that one arm be designated the 'plus' (P) arm and drawn upwards in the idiogram and the other arm be designated 'minus' (M) and drawn downwards with the centromere at position 0. We have adopted these designations to our figures with the previously designated 'short' and 1β

arms becoming the 'plus' arms and the 'long' and 1α arms becoming the 'minus' arms.

The wheat-barley addition lines have provided evidence that barley chromosomes 1, 2, 3, 4, 5, 6 and 7 are homoeologous to wheat chromosomes and are sometimes designated 7H, 2H, 3H, 4H, 1H, 6H and 5H, respectively. This conclusion is supported by the occurrence of similar morphological, biochemical and molecular markers on the homoeologous chromosomes and the ability of the barley chromosomes to substitute for the equivalent wheat chromosomes in substitution lines (reviewed in Shepherd and Islam 1992). In this report we use both chromosome designations since some reports use one or the other terminology.

The genetic length of the barley chromosome maps, as defined by telomeric markers on both chromosome arms, has not yet been determined. Therefore our conclusions about the genetic lengths of the chromosomes are preliminary. The RFLP and other markers located on the maps reviewed herein are listed in Tables 1–7, arranged by the crosses used for the mapping and starting from the most terminal marker on the plus arm and ending with the most terminal marker on the minus arm within each cross. The number of markers mapped and the genetic distance in centiMorgans (cM) for each chromosome within each of the major populations used for mapping are summarized in Table 8. These data show that chromosomes 1, 2, 3 and 7 are the longest and chromosomes 4, 5 and 6, somewhat shorter in terms of their genetic or recombination lengths. This corresponds reasonably well with their physical lengths, although it is well established that within any one chromosome there is poor correspondence between the physical and genetic distances (reviewed in Ramage 1985; Kleinhofs et al. 1988). The maps developed by the NABGMP using the Steptoe by Morex cross are presented in Figs. 1–5. The NABGMP maps were aligned with other maps where possible (Figs. 1, 2, 4 and 5). Readers are referred to the original publications for the other barley RFLP maps (Heun et al. 1991; Graner et al. 1991; Hinze et al. 1991).

4.1. Chromosome 1 (7H)

A total of 126 markers have been assigned to chromosome 1 RFLP maps (Table 1). Since 16 markers appear on more than one map, this represents 110 probable unique markers available for chromosome 1. No marker appears on all five maps. The waxy endosperm locus, designated *Glx* (also *wx*), encoding starch synthase (Rohde et al. 1988) is located on three of the RFLP maps (S/M; I/F; S/W). This locus is also found on the classical barley maps (Sogaard and von Wettstein-Knowles 1987), thus providing a starting point for merging all maps. Eight markers identifying the same locus are shared between the S/M and P/N maps, 4 markers between the S/M and S/W maps, and 4 markers between the V/H.s. and I/F maps. The number of shared markers allows a reasonable alignment of the S/M map with P/N and S/W maps (Fig. 1). The I/F and V/H.s. maps have been previously aligned (Graner et al. 1991) and do not share enough markers with the other maps for further alignment at this time.

Table 1. Markers mapped to barley chromosome 1 (7H).^a

Designation	Description	Rating ^b	Source	Reference
Steptoe by Morex Cross (S/M)				
Tel1P	barley sub-telomeric gDNA	P	NABGMP ^c	Kilian (unpub.)
Plc	plastocyanin	E	K. Gausing	Nielsen and Gausing (1987)
ABR303	CC GC CT AG TC (primer)	F	NABGMP	Kleinhofs (1992)
BCD129	barley cDNA	G	M. Sorrells	Heun et al. (1991)
iEst5	esterase (isozyme)	G	P. Hayes	Nielsen and Johansen (1986)
Glx	starch synthase(waxy)	G	W. Rohde	Rohde et al. (1988)
Prx1A	seed perox B1 cDNA	G	S. Rasmussen	Rasmussen et al. (1991)
His3	histone 3	F	A. Brandt	Chojecki (1986)
WG789	wheat gDNA	F	M. Sorrells	Heun et al. (1991)
ABC151	barley malt cDNA	G	NABGMP	Kleinhofs (1992)
WG834	wheat gDNA	G	M. Sorrells	Heun et al. (1991)
CDO475	oat cDNA	E	" "	" "
ABG380	barley gDNA	G	NABGMP	Kleinhofs (1992)
ABC158	barley malt cDNA	G	" "	" "
ksuA1A	<i>T. tauschii</i> gDNA	G	B. Gill	Gill et al. (1991)
ABC154	barley malt cDNA	F	NABGMP	Kleinhofs (1992)
Brz	glucosyl transferase (bronze)	G	R. Wise	Wise et al. (1990)
ABC465	barley leaf cDNA	P	NABGMP	Kleinhofs (1992)
ABC156D	barley malt cDNA	F	" "	" "
Adh7	alcohol dehydrogenase	P	A. Good	Good et al. (1988)
ABR329	TG TT CC AC GG (primer)	F	NABGMP	Kleinhofs (1992)
ABC455	barley leaf cDNA	E	" "	" "
ABG476	barley gDNA	G	" "	" "
WG719	wheat gDNA	F	M. Sorrells	Heun et al. (1991)
CDO673	oat cDNA	E	" "	" "
Amy2	α -amylase	G	J. Rogers	Khursheed and Rogers (1988)
Ubi1	ubiquitin	P	K. Gausing	Gausing and Barkardottir (1986)
ABC310B	barley malt cDNA	F	NABGMP	Kleinhofs (1992)
ABC305	barley malt cDNA	G	" "	" "
Pgk2B	p-glycerate kinase, cyto.	F	S. Chao	Chao et al. (1989)
PSR129	wheat leaf cDNA	F	M. Gale	Chao et al. (1988)
ABG461	barley gDNA	G	NABGMP	Kleinhofs (1992)
WG420	wheat gDNA	G	M. Sorrells	Heun et al. (1991)
WG380	" "	G	" "	" "
ksuD14	<i>T. tauschii</i> gDNA	G	B. Gill	Gill et al. (1991)
Dor4B	dormin group 4	F	K. Walker-Simmons	Morris et al. (1991)
PSR106B	wheat leaf cDNA	G	M. Gale	Chao et al. (1988)
Chi1	chitinase, 26 kDa	F	R. Leah	Leah et al. (1991)
BCD298B	barley cDNA	G	M. Sorrells	Heun et al. (1991)
Tha	thaumatin-like protein	E	R. Dudler	Rebmann et al. (1991)

Table 1. Continued.

Designation	Description	Rating ^b	Source	Reference
Proctor by Nudinka Cross (P/N)				
CDO545	oat cDNA	F	M. Sorrells	Heun et al. (1991)
CDO420A	"	P	"	"
BCD129	barley cDNA	G	"	"
BCD130	"	G	"	"
WG834	wheat gDNA	G	"	"
WG789	"	F	"	"
CDO475	oat cDNA	E	"	"
CDO36	"	F	"	"
CDO348	"	F	"	"
CDO771B	"	F	"	"
WG719	wheat gDNA	F	"	"
BCD205	barley cDNA	E	"	"
CDO464	oat cDNA	F	"	"
BCD421	barley cDNA	F	"	"
CDO687	oat cDNA	F	"	"
CDO358	"	F	"	"
WG669	wheat gDNA	E	"	"
CDO689	oat cDNA	P	"	"
CDO595	"	G	"	"
CDO673	"	F	"	"
n	naked/covered seeds	E	"	"
BG143	barley gDNA	P	N. Lapitan	"
BCD351A	barley cDNA	G	M. Sorrells	"
WG380B	wheat gDNA	G	"	"
WG686	"	G	"	"
WG338B	"	G	"	"
WG420	"	G	"	"
WG380A	"	G	"	"
WG420B	"	G	"	"
CDO347	oat cDNA	P	"	"
BCD512	barley cDNA	F	"	"
BG141	barley gDNA	F	N. Lapitan	"
Vada by <i>Hordeum vulgare</i> subsp. <i>spontaneum</i> line 1B-87 Cross (V/H.s.)				
MWG35	barley gDNA	?	A. Graner	Graner et al. (1991)
MWG47	"	?	"	"
MWG530	"	?	"	"
MWG622b	"	?	"	"
MWG4a	"	?	"	"
MWG527	"	?	"	"
MWG18	"	?	"	"
MWG606	"	?	"	"
MWG39	"	?	"	"
MWG3	"	?	"	"
MWG66	"	?	"	"
MWG710a	"	?	"	"
MWG626	"	?	"	"
MWG622a	"	?	"	"
MWG53	"	?	"	"

Table 1 Continued

Designation	Description	Rating ^b	Source	Reference
MWG633	„	?	„	„
MWG70	„	?	„	„
MWG24	„	?	„	„
MWG86	„	?	„	„
Igri by Franka Cross (I/F)				
MWG555a	barley gDNA	?	A Graner	Graner et al (1991)
MWG530	„	?	„	„
Est5	esterase 5 (isozyme)	?	„	Kahler and Allard (1970)
pWx27	starch synthase(waxy)	G	W Rohde	Rohde et al (1988)
cMWG703	barley leaf cDNA	?	A Graner	Graner et al (1991)
MWG564	barley gDNA	?	„	„
MWG527	„	?	„	„
MWG89	„	?	„	„
cMWG721	barley leaf cDNA	?	„	„
cMWG773	„	?	„	„
MWG622	barley gDNA	?	„	„
cMWG739	barley leaf cDNA	?	„	„
cMWG725	„	?	„	„
cMWG714	„	?	„	„
cMWG705	„	?	„	„
cMWG649a	„	?	„	„
MWG626	barley gDNA	?	„	„
cMWG741	barley leaf cDNA	?	„	„
MWG511	barley gDNA	?	„	„
cMWG704	barley leaf cDNA	?	„	„
cMWG729	„	?	„	„
cMWG696	„	?	„	„
MWG528	barley gDNA	?	„	„
MWG599	„	?	„	„
MWG633	„	?	„	„
MWG539	„	?	„	„
Steptoe mutant Az12 by Winer Cross (S/W)				
PSR119	wheat leaf cDNA	G	M Gale	Kilian (unpub)
Glx1	starch synthase	G	W Rohde	Rohde et al (1988)
Brz1	glucosyl transferase (bronze)	G	R Wise	Wise et al (1990)
PSR108	wheat leaf cDNA	E	M Gale	Kilian (unpub)
PSR103	„	G	„	„
PSR105	„	G	„	„
Amy2	α -amylase	G	J Rogers	Khursheed and Rogers (1988)
PSR117	wheat leaf cDNA	G	M Gale	Kilian (unpub)
PSR129	„	G	„	„
PSR121	„	G	„	„

^a Markers are arranged in order from the most distal on the plus arm to the most distal on the minus arm within each population used to construct the maps

^b E - excellent, G - good, F - fair, P - poor

^c NABGMP - North American Barley Genome Mapping Project

CHROMOSOME 1 (7H)

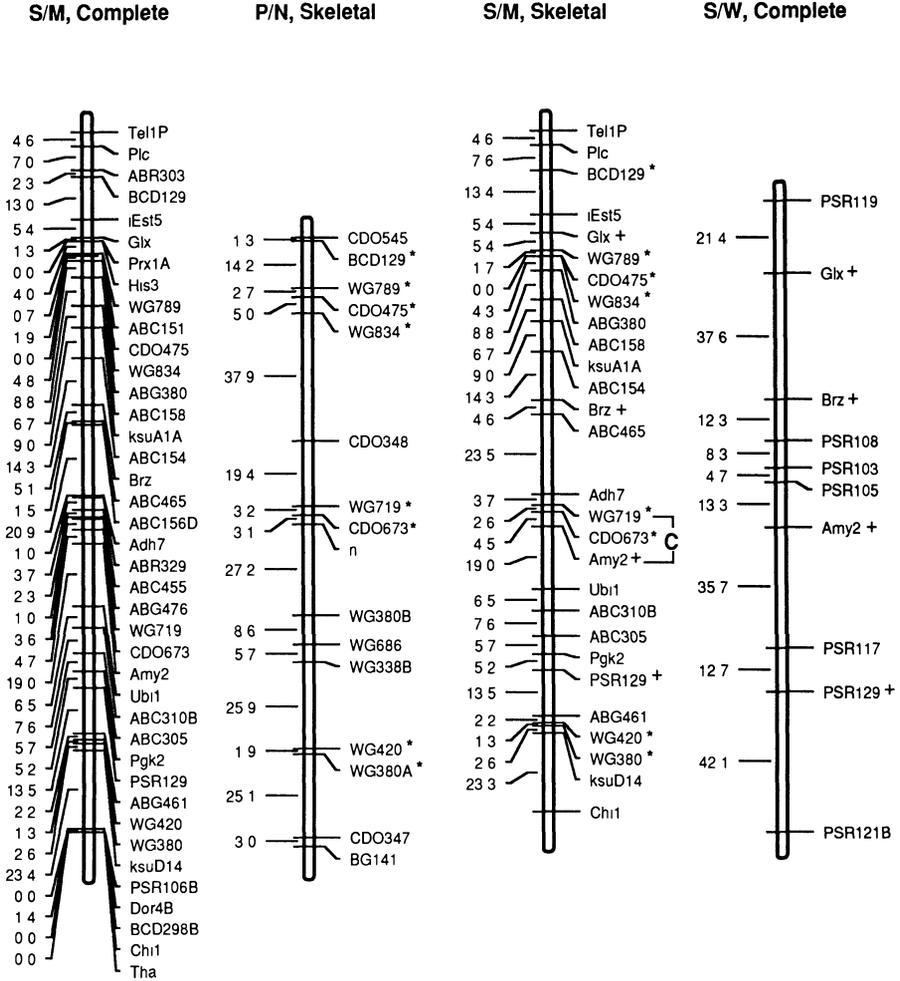


Fig 1 Chromosome 1 (7H) map, based on the Steptoe by Morex (S/M) cross, is compared to those developed with the Proctor by Nudinka (P/N) cross and the Steptoe mutant Az12 by Winer (S/W) cross. Skeletal maps are presented to facilitate comparison of marker order and distances. Markers identifying the same loci in the S/M and P/N maps and S/M and S/W maps are marked with a '*' and '+', respectively. The approximate centromere location is marked with a 'C' and a bracket identifying the two most proximal markers tested on each arm.

Chromosome 1 centromere was located between bronze (*Brz*) and α -amylase 2 (*Amy 2*) on the S/M map. WG719 was recently located to chromosome 1 plus arm (Sorrells, pers. comm.), thus the centromere location is probably between WG719 and *Amy 2* (Fig. 1). Recently, the NABGMP has succeeded in cloning

several putative subtelomeric regions. A presumed chromosome 1 plus arm telomeric polymorphism maps ca. 5 cM distal to plastocyanin (*Plc*) (Kilian, unpub.). This suggests that the map of the plus arm of chromosome 1 is essentially full length. The degree of completeness of the minus arm can not be estimated at this time. An unusual clustering of 5 unrelated probes showing no recombination is observed at the terminus of the chromosome 1 minus arm in the S/M map (Fig. 1). Similar clustering of markers is seen on the minus arm of chromosome 1 in the I/F map, but this cluster maps somewhat more proximal on the chromosome 1 minus arm (Graner et al. 1991). This phenomenon is not seen in the other maps and may represent a chromosomal rearrangement, such as an inversion or translocation, which could lower recombination.

4.2. Chromosome 2 (2H)

A total of 125 markers have been placed on the chromosome 2 RFLP maps (Table 2). Since 15 of the 125 markers probably identify the same locus on two maps, there are 110 unique markers available for chromosome 2. The S/M map shares 10 markers with the P/N map allowing partial alignment of the two maps (Fig. 2). Two additional markers, BCD351 and CDO 474, map to different locations in the P/N and S/M maps. The CDO474 probe probably hybridizes to a different polymorphic band in the S/M population and therefore has been designated CDO474C to differentiate it from the CDO474 locus on chromosome 3 of the P/N map. The BCD351 probe appears to detect cross-hybridizing loci scattered throughout the barley genome. This unusual probe and the loci it detects will be discussed in more detail later. The I/F and V/H.s. maps share 5 markers and have been previously aligned (Graner et al. 1991). The only marker relating these maps to the classical map is the 2-row versus 6-row locus *hex-v* (also *V*) located on the I/F map.

Table 2. Markers mapped to barley chromosome 2 (2H).^a

Designation	Description	Rating ^b	Source	Reference
Stephoe by Morex Cross (S/M)				
Tel2P	subtelomeric barley gDNA	F	NABGMP ^c	Kilian (unpub.)
Chs1B	chalcone synthase	P	W. Rohde	Rohde et al. (1991)
Chs1A	"	P	"	"
WG516	wheat gDNA	F	M. Sorrells	Heun et al. (1991)
RbcS	RuBP C/O small subunit	E	K. Gausing	Barkardottir et al. (1987)
BCD351F	barley cDNA	F	M. Sorrells	Heun et al. (1991)
ABC156A	barley malt cDNA	F	NABGMP	Kleinhofs (1992)
ABG358	barley gDNA	E	"	"
CDO64	oat cDNA	F	M. Sorrells	Heun et al. (1991)
ABG459	barley gDNA	E	NABGMP	Kleinhofs (1992)
Pox	pathogen-induced peroxidase	E	R. Dudler	Hertig et al. (1991)

Table 2 Continued

Designation	Description	Rating ^b	Source	Reference
ABC454	barley leaf cDNA	G	NABGMP	Kleinhofs (1992)
Adh8	alcohol deh , cDNA	P	A Good	Good et al (1988)
CDO537	oat cDNA	G	M Sorrells	Heun et al (1991)
ABC468	barley leaf cDNA	F	NABGMP	Kleinhofs (1992)
Bmy2	β -amylase	G	M Kreis	Kreis et al (1987)
ABC306	barley leaf cDNA	G	NABGMP	Kleinhofs (1992)
ABC162	barley malt cDNA	E	"	"
ABG356	barley gDNA	E	"	"
WG996	wheat gDNA	G	M Sorrells	Heun et al (1991)
CDO588	oat cDNA	G	"	"
ABC451	barley leaf cDNA	G	NABGMP	Kleinhofs (1992)
CDO474C	oat cDNA	F	M Sorrells	Heun et al (1991)
ABC152D	barley malt cDNA	G	NABGMP	Kleinhofs (1992)
ksuF15	<i>T tauschu</i> gDNA	E	B S Gill	Gill et al (1991)
ksuD22	"	E	"	"
CDO373	oat cDNA	G	M Sorrells	Heun et al (1991)
Gln2	glutamine synthase, chlor	E	B G Forde	Freeman et al (1990)
ABC157	barley malt cDNA	E	NABGMP	Kleinhofs (1992)
ABC153	"	G	"	"
ABC165	"	G	"	"
Pcr1	protochlorophyllide reductase	G	K Apel	Schulz et al (1989)
BCD410	barley cDNA	G	M Sorrells	Heun et al (1991)
Prx2	leaf peroxidase, cDNA	G	S Rasmussen	unpub
BG123	barley gDNA	F	N Lapitan	Heun et al (1991)
WG645	wheat gDNA	G	M Sorrells	"
Proctor by Nudinka Cross (P/N)				
BG140B	barley gDNA	F	N Lapitan	Heun et al (1991)
ksuD18	<i>T tauschu</i> gDNA	G	B S Gill	Gill et al (1991)
CDO57	oat cDNA	G	M Sorrells	Heun et al (1991)
BCD175	barley cDNA	F	"	"
WG516	wheat gDNA	G	"	"
BCD221B	barley cDNA	F	"	"
WG222A	wheat gDNA	G	"	"
CDO64	oat cDNA	G	"	"
CDO665	"	G	"	"
BCD351B	barley cDNA	G	"	"
CDO370	oat cDNA	F	"	"
WG405A	wheat gDNA	G	"	"
CDO770	oat cDNA	F	"	"
ksuE12	<i>T tauschu</i> gDNA	G	B S Gill	Gill et al (1991)
CDO675A	oat cDNA	F	M Sorrells	Heun et al (1991)
CDO537	"	G	"	"
WG996	wheat gDNA	G	"	"
CDO667	oat cDNA	G	"	"
CDO366	"	G	"	"
BCD111	barley cDNA	E	"	"
WG223	wheat gDNA	F	"	"

Table 2 Continued

Designation	Description	Rating ^b	Source	Reference
CDO588	oat cDNA	G	"	"
BCD355	barley cDNA	G	"	"
BCD334	"	F	"	"
CDO373	oat cDNA	F	"	"
BCD453	barley cDNA	F	"	"
CDO680	oat cDNA	G	"	"
ksuF15	<i>T. tauschii</i> gDNA	G	B S Gill	Gill et al (1991)
BCD266	barley cDNA	G	M Sorrells	Heun et al (1991)
ksuF41	<i>T. tauschii</i> gDNA	G	B S Gill	Gill et al (1991)
ksuG43	"	E	"	"
CDO678	oat cDNA	G	M Sorrells	Heun et al (1991)
WG338A	wheat gDNA	G	"	"
BCD410	barley cDNA	P	"	"
ksuH16	<i>T. tauschii</i> gDNA	G	B S Gill	Gill et al (1991)
WG645	wheat gDNA	G	M Sorrells	Heun et al (1991)
BCD339C	barley cDNA	F	"	"
BG123	barley gDNA	G	N Lapitan	"

Vada by *Hordeum vulgare* subsp. *spontaneum* line 1B-87 Cross (V/H.s.)

MWG23	barley gDNA	?	A Graner	Graner et al (1991)
cMWG682	barley leaf cDNA	?	"	"
MWG83	barley gDNA	?	"	"
MWG48	"	?	"	"
MWG84b	"	?	"	"
MWG52	"	?	"	"
MWG64	"	?	"	"
cMWG737	barley leaf cDNA	?	"	"
MWG9	barley gDNA	?	"	"
MWG65	"	?	"	"
MWG81	"	?	"	"
MWG71	"	?	"	"
MWG84a	"	?	"	"
cMWG657	barley leaf cDNA	?	"	"
cMWG656	"	?	"	"
cMWG702	"	?	"	"
cMWG726	"	?	"	"
MWG5	barley gDNA	?	"	"
cMWG738	barley leaf cDNA	?	"	"
MWG1	barley gDNA	?	"	"
MWG503	"	?	"	"
MWG87	"	?	"	"
cMWG694	barley leaf cDNA	?	"	"
cMWG671	"	?	"	"
MWG82	barley gDNA	?	"	"
MWG46	"	?	"	"
MWG17	"	?	"	"
cMWG660	barley leaf cDNA	?	"	"
MWG80	barley gDNA	?	"	"

Table 2 Continued

Designation	Description	Rating ^b	Source	Reference
Igri by Franka Cross (I/F)				
cMWG684b	barley leaf cDNA	?	A Graner	Graner et al (1991)
cMWG682	„	?	„	„
cMWG655	„	?	„	„
MWG64	barley gDNA	?	„	„
MWG578	„	?	„	„
MWG553b	„	?	„	„
MWG520	„	?	„	„
cMWG663	barley leaf cDNA	?	„	„
cMWG646a	„	?	„	„
cMWG647	„	?	„	„
MWG65	barley gDNA	?	„	„
MWG557	„	?	„	„
cMWG658	barley leaf cDNA	?	„	„
hex-v	2-row vs 6-row, morph	?	„	„
cMWG699	barley leaf cDNA	?	„	„
MWG801	barley gDNA	?	„	„
MWG503	„	?	„	„
MWG581	„	?	„	„
cMWG720	barley leaf cDNA	?	„	„
cMWG660	„	?	„	„
MWG90	barley gDNA	?	„	„
MWG636	„	?	„	„

^a Markers are arranged in order from the most distal on the plus arm to the most distal on the minus arm within each population used to construct the maps

^b E – excellent, G – good, F – fair, P – poor

^c NABGMP – North American Barley Genome Mapping Project

The chromosome 2 centromere appears to be located between the *Bmy2* and *ABG356* loci (Fig. 2). A putative subtelomeric sequence has been isolated and mapped on chromosome 2P arm, indicating that this arm of chromosome 2 is probably full length. The extent of coverage of the minus arm of this chromosome is not possible to determine at this time.

4.3. Chromosome 3 (3H)

A total of 86 markers have been placed on the chromosome 3 RFLP maps (Table 3). Since 4 markers identify a common locus on two maps, there are 82 probable unique markers available for chromosome 3. The V/H.s. map shares three markers with the I/F map (Graner et al. 1991) and the S/M map shares one marker with the P/N map. Since it was not possible to align the maps, only the NABGMP map developed with the S/M cross is shown (Fig. 3). No markers on the molecular maps are in common with the classical map in spite of the fact that the chromosome 3 classical map is much better than those for chromosomes 1 and 2 (Sogaard and von Wettstein-Knowles 1987).

CHROMOSOME 2 (2H)

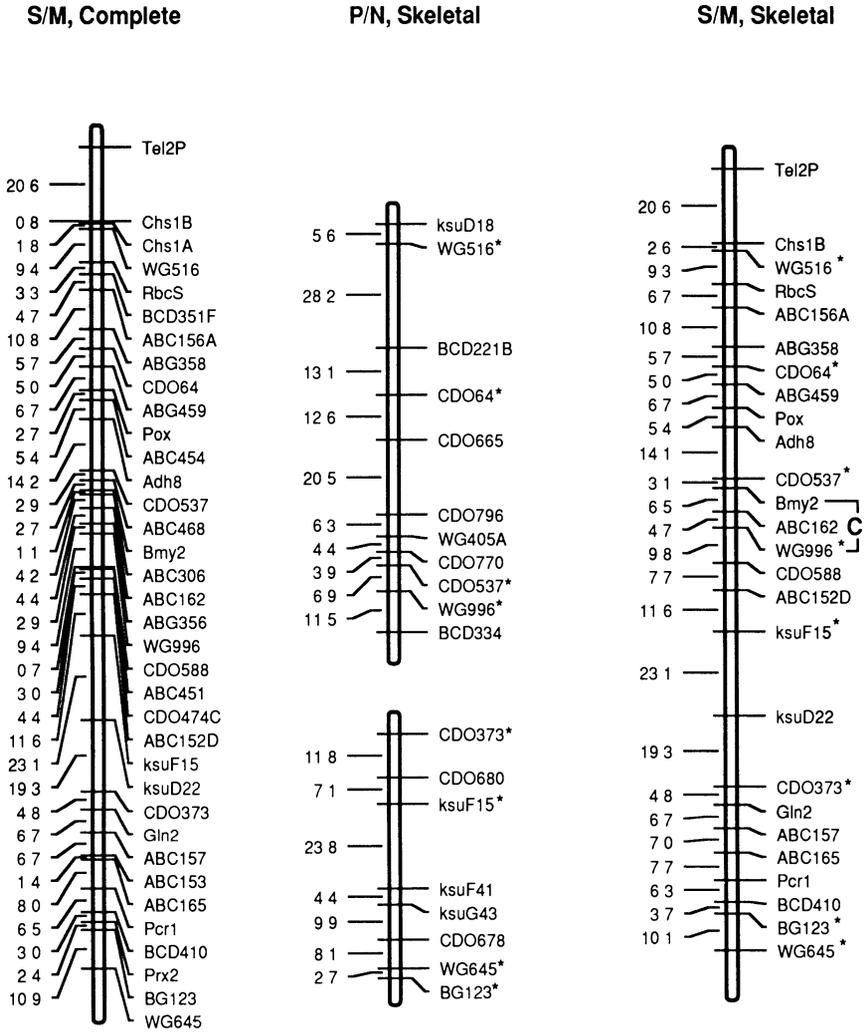


Fig 2 Chromosome 2 (2H) map, based on the Steptoe by Morex (S/M) cross is compared to the map developed with the Proctor by Nudinka (P/N) cross. Skeletal maps are presented to facilitate comparison of marker order and distances. Markers identifying the same loci in the S/M and P/N maps are marked with a '*'. The approximate centromere location is marked with a 'C' and a bracket identifying the two most proximal markers tested on each arm.

Table 3 Markers mapped to barley chromosome 3 (3H) ^a

Designation	Description	Rating ^b	Source	Reference
Stepptoe by Morex Cross (S/M)				
ABG460	barley gDNA	E	NABGMP ^c	Kleinhofs (1992)
ABG471	"	G	"	"
ABG462	"	G	"	"
ABC156C	barley malt cDNA	F	"	"
ksuF2B	<i>T tauschii</i> gDNA	P	B S Gill	Gill et al (1991)
Adh5	alcohol deh , cDNA	P	A Good	Good et al (1988)
BCD828	barley cDNA	E	M Sorrells	Heun et al (1991)
ksuA3C	<i>T tauschii</i> gDNA	G	B S Gill	Gill et al (1991)
Dor4A	dormin, group 4	P	K Walker-Simmons	Morris et al (1991)
PSR156	wheat leaf cDNA	G	M Gale	Chao et al (1988)
ABG377	barley gDNA	F	NABGMP	Kleinhofs (1992)
ABG453	"	G	"	"
ABC307B	barley cDNA	G	W Rohde	unpub
CDO113A	oat cDNA	P	M Sorrells	Heun et al (1991)
His4B	histone 4	P	M Iwabuchi	Tabata et al (1983)
ABG4	barley gDNA	E	NABGMP	Kleinhofs (1992)
WG110	wheat gDNA	G	M Sorrells	Heun et al (1991)
mPub	pubescent leaf	G	J Franckowiak	unpub
ABC161	barley malt cDNA	G	NABGMP	Kleinhofs (1992)
Glb4	(1-3)- β -glucanase	G	R Leah	Leah et al (1991)
Glb3	"	G	"	"
1Bgl	β -galactosidase, isozyme	-	D Hoffman	-
Prx1B	peroxidase B1, seed	P	S Rasmussen	Rasmussen et al (1991)
ABC166	barley malt cDNA	G	NABGMP	Kleinhofs (1992)
Proctor by Nudinka Cross (P/N)				
CDO1174	oat cDNA	F	M Sorrells	Heun et al (1991)
CDO395	"	E	"	"
CDO684	"	G	"	"
WG178	wheat gDNA	E	"	"
BCD339A	barley cDNA	F	"	"
WG940	wheat gDNA	G	"	"
WG405B	"	G	"	"
CDO419B	oat cDNA	F	"	"
WG110	wheat gDNA	G	"	"
BCD131	barley cDNA	F	"	"
CDO105	oat cDNA	F	"	"
CDO474	"	G	"	"
BG131	barley gDNA	E	N Lapitan	"
CDO394A	oat cDNA	P	M Sorrells	"
Vada by <i>Hordeum vulgare</i> subsp. <i>spontaneum</i> line 1B-87 Cross (V/H.s.)				
cMWG691a	barley cDNA	?	A Graner	Graner et al (1991)
MWG46	barley gDNA	?	"	"
MWG33	"	?	"	"
MWG31	"	?	"	"

Table 3 Continued

Designation	Description	Rating ^b	Source	Reference
MWG584	..	?
MWG14	..	?
MWG22	..	?
cMWG687	barley cDNA	?
MWG8	barley gDNA	?
MWG61	..	?
MWG631	..	?
MWG595	..	?
MWG823	..	?
cMWG680	barley cDNA	?
MWG43	barley gDNA	?
MWG67	..	?
MWG818	..	?
cMWG661	barley cDNA	?
MWG54	barley gDNA	?
MWG91	..	?
MWG44	..	?
MWG806	..	?
MWG30	..	?
MWG73	..	?
MWG69	..	?
MWG570	..	?
MWG6	..	?
MWG41	..	?
MWG11a	..	?
MWG630	..	?
Igri by Franka Cross (I/F)				
MWG677	barley gDNA	?	A Graner	Graner et al (1991)
MWG584	..	?
MWG595	..	?
MWG561a	..	?
MWG802	..	?
MWG582	..	?
cMWG680	barley cDNA	?
MWG812	barley gDNA	?
MWG555b	..	?
MWG571	..	?
MWG632a	..	?
MWG570	..	?
cMWG693	barley cDNA	?
MWG546	barley gDNA	?
Est1	esterase, isozyme	?
MWG85	barley gDNA	?
MWG10b	..	?
cMWG691b	barley cDNA	?

^a Markers are arranged in order from the most distal on the plus arm to the most distal on the minus arm within each population used to construct the maps

^b E - excellent, G - good, F - fair, P - poor

^c NABGMP - North American Barley Genome Mapping Project

CHROMOSOME 3 (3H)

CHROMOSOME 4 (4H)

CHROMOSOME 5 (1H)

S/M, Complete

S/M, Complete

S/M, Complete

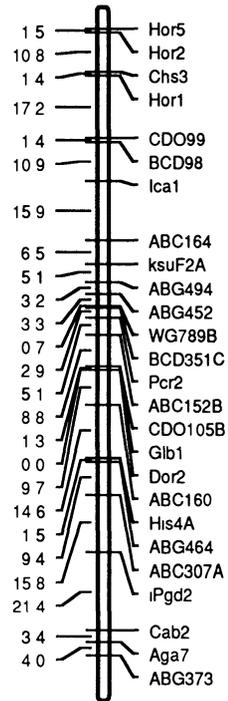
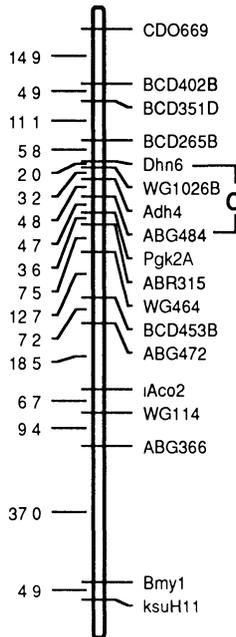
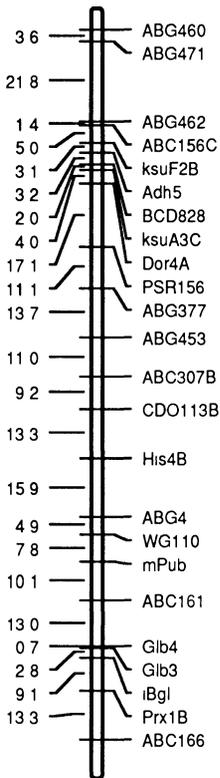


Fig 3 Chromosome 3 (3H), 4 (4H) and 5 (1H) maps presented are based on the doubled haploid population derived from the Steptoe by Morex (S/M) cross. The approximate centromere location is marked with a 'C' and a bracket identifying the two most proximal markers tested on each arm of chromosome 4. The centromere locations for chromosomes 3 and 5 have not been identified.

The chromosome 3 centromere has not been localized and no telomeric regions have been mapped. Thus, it is difficult to draw any conclusions about the genetic length of the chromosome 3 map.

4.4. Chromosome 4 (4H)

A total of 63 markers have been placed on the chromosome 4 RFLP maps (Table 4). Since 7 markers identify probable identical loci on two maps, there are 56 unique markers available for chromosome 4. Since there are insufficient

common markers to align the various maps, only the NABGMP map developed with the S/M cross is presented (Fig. 3). Seven probes used to construct the P/N map were also placed on the S/M map, but only three appear to be located in similar positions. The others, BCD265, BCD402, BCD453, and WG1026, appear to involve different polymorphic bands that map at different locations in the barley genome. This clearly illustrates the potential problems that can be encountered in transferring probes that hybridize to more than one band among different laboratories and to different crosses. The map developed around the *ml-o* locus (Hinze et al. 1991) can be aligned with the S/M map due to the fortuitous inclusion of the *Bmy1* marker in both maps. Three markers mapped to the I/F and V/H.s. maps appear to identify the same loci (Graner et al. 1991).

Table 4 Markers mapped to barley chromosome 4 (4H) ^a

Designation	Description	Rating ^b	Source	Reference
Steptoe by Morex Cross (S/M)				
CDO669	oat cDNA	F	M Sorrells	Heun et al (1991)
BCD402	barley cDNA	P	"	"
BCD351D	"	F	"	"
BCD265	"	G	"	"
Dhn6	dehydrin	E	T Close	Close and Chandler (1990)
WG1026B	wheat gDNA	G	M Sorrells	Heun et al (1991)
Adh4	alcohol deh , cDNA	P	A Good	Good et al (1988)
ABG484	barley gDNA	G	NABGMP ^c	Kleinhofs (1992)
Pgk2A	P-glycerate kinase, cytosol	F	S Chao	Chao et al (1989)
ABR315	CT GC TT AG GG primer	F	NABGMP	Kleinhofs (1992)
WG464	wheat gDNA	G	M Sorrells	Heun et al (1991)
BCD453B	barley cDNA	P	"	"
ABG472	barley gDNA	E	NABGMP	Kleinhofs (1992)
1Aco2	aconitase, isozyme	-	D Hoffman	unpub
WG114	wheat gDNA	G	M Sorrells	Heun et al (1991)
ABG366	barley gDNA	E	NABGMP	Kleinhofs (1992)
Bmy1	β -amylase	G	M Kreis	Kreis et al (1987)
ksuH11	<i>T tauschu</i> gDNA	F	B S Gill	Gill et al (1991)
Proctor by Nudinka Cross (P/N)				
WG622	wheat gDNA	G	M Sorrells	Heun et al (1991)
CDO669	oat cDNA	G	"	"
CDO541	"	E	"	"
WG181	wheat gDNA	G	"	"
WG464	"	E	"	"
WG180B	"	G	"	"
CDO586	oat cDNA	G	"	"
CDO795	"	G	"	"
WG232	wheat gDNA	E	"	"
CDO650	oat cDNA	G	"	"

Table 4. Continued.

Designation	Description	Rating ^b	Source	Reference
BG125	barley gDNA	G	"	"
WG114	wheat gDNA	G	"	"
CDO63	oat cDNA	F	"	"
BCD402	barley cDNA	G	"	"
WG199	wheat gDNA	G	"	"
CDO465	oat cDNA	P	"	"
Vada by <i>Hordeum vulgare</i> subsp. <i>spontaneum</i> line 1B-87 Cross (V/H.s.)				
MWG634	barley gDNA	?	A. Graner	Graner et al. (1991)
MWG77	"	?	"	"
MWG11b	"	?	"	"
MWG611	"	?	"	"
MWG32	"	?	"	"
MWG29	"	?	"	"
MWG27	"	?	"	"
MWG58	"	?	"	"
cMWG677	barley cDNA	?	"	"
MWG57	barley gDNA	?	"	"
MWG542	"	?	"	"
cMWG708	barley cDNA	?	"	"
MWG42	barley gDNA	?	"	"
MWG88	"	?	"	"
cMWG710c	barley cDNA	?	"	"
MWG51	barley gDNA	?	"	"
MWG616	"	?	"	"
MWG517	"	?	"	"
Igri by Franka Cross (I/F)				
MWG635	barley gDNA	?	A. Graner	Graner et al. (1991)
MWG58	"	?	"	"
MWG57	"	?	"	"
MWG542	"	?	"	"
Aramir by <i>Hordeum vulgare</i> subsp. <i>spontaneum</i> (A/H.s.)				
b4-104/1	barley gDNA	?	K. Hinze	Hinze et al. (1991)
bBE54a	"	?	"	"
bAN61	"	?	"	"
bAP91	"	?	"	"
bAL88/2	"	?	"	"
bAO11	"	?	"	"
Bmy1	β -amylase	?	M. Gale	Kreis et al. (1987)

^a Markers are arranged in order from the most distal on the plus arm to the most distal on the minus arm within each population used to construct the maps.

^b E – excellent; G – good; F – fair; P – poor.

^c NABGMP – North American Barley Genome Mapping Project.

The chromosome 4 centromere has been localized between the markers *Dhn6* and ABG484 on the S/M map (Fig. 3). The full genetic length of chromosome 4 can not be determined at this time due to absence of telomeric markers. The paucity of markers in the *ml-o* region of chromosome 4 was noted (Hinze et al. 1991). This may extend to the whole chromosome as indicated by the low number of markers on all of the chromosome 4 RFLP maps. The powdery mildew resistance locus *ml-o* is the only marker shared between the molecular and classical maps.

4.5. Chromosome 5 (1H)

A total of 78 markers have been placed on the chromosome 5 RFLP maps (Table 5). Since 2 markers identify probable identical loci on three maps and 6 markers on two maps, there are 68 unique markers available for chromosome 5. Since there are insufficient common markers to align the RFLP maps, only the NABGMP map based on the S/M cross is presented (Fig. 3). Three loci, *Hor2*, *Ml-a*, and *Hor1* (distal to proximal order), provide a useful starting point for future alignment of the RFLP maps among themselves and with the classical map. The *Hor1* and *Hor2* loci have been mapped on the S/M, V/H.s., and I/F molecular maps where they define the distal region of the chromosome 5P arm. The *Ml-a* locus has been mapped on the P/N and I/F maps. However, due to its location between the *Hor1* and *Hor2* loci, it can easily be approximated on the other maps. A single marker, *ksu DG D14A*, extends ca. 20 cM distal from the *Ml-a* locus on the P/N map. The classical map extends about 30 cM distal from the *Hor2* locus to another mildew resistance locus *Ml-at* (Sogaard and von Wettstein-Knowles 1987), suggesting that additional distal markers should be found for the molecular maps of chromosome 5P.

Table 5 Markers mapped to barley chromosome 5 (1H) ^a

Designation	Description	Rating ^b	Source	Reference
Stephoe by Morex Cross (S/M)				
Hor5	γ -hordeins	F	A Brandt	Cameron-Mills and Brandt (1988)
Hor2	B-hordeins	G	P R Shewry	Forde et al (1985)
Chs3	chalcone synthase	P	W Rohde	Rohde et al (1991)
Hor1	C-hordeins	G	P R Shewry	Forde et al (1985)
CDO99	oat cDNA	E	M Sorrells	Heun et al (1991)
BCD98	barley cDNA	F	"	"
Ica1	chymotrypsin inhibitor	G	P R Shewry	Williamson et al (1988)
ABC164	barley malt cDNA	F	NABGMP ^c	Kleinshofs (1992)
ksuF2A	<i>T tauschii</i> gDNA	P	B S Gill	Gill et al (1991)
ABG494	barley gDNA	E	NABGMP	Kleinshofs (1992)
ABG452	"	G	"	"
WG789B	wheat gDNA	F	M Sorrells	Heun et al (1991)

Table 5. Continued.

Designation	Description	Rating ^b	Source	Reference
BCD351C	barley cDNA	F	"	"
Pcr2	protochlorophyllide red.	F	K. Apel	Schultz et al. (1989)
ABC152B	barley malt cDNA	G	NABGMP	Kleinhofs (1992)
CDO105	oat cDNA	F	M. Sorrells	Heun et al. (1991)
Glb1	(1-3;1-4) β -glucanase	G	J. Litts	Litts et al. (1990)
Dor2	dormin, group 2	G	K. Walker-Simmons	Morris et al. (1991)
ABC160	barley malt cDNA	E	NABGMP	Kleinhofs (1992)
His4A	histone 4	P	M. Iwabuchi	Tabata et al. (1983)
ABG464	barley gDNA	G	NABGMP	Kleinhofs (1992)
ABC307A	barley cDNA	G	W. Rohde	unpub.
iPg2	phosphogluco deh., isozyme	E	P. Hayes	Nielsen and Johansen (1986)
Cab2	chlor a/b binding protein	F	K. Gausing	Barkardottir et al. (1987)
Aga7	ADP-gluc ppase, endosperm	G	M.R Olive	Olive et al. (1989)
ABG373	barley gDNA	E	NABGMP	Kleinhofs (1992)
Proctor by Nudinka Cross (P/N)				
ksuD14A	<i>T. tauschii</i> gDNA	G	B.S. Gill	Gill et al. (1991)
Mla12	powdery mildew resistance	-	M. Sorrells	Heun et al. (1991)
22Epr8	barley promoter	-	O.-A. Olsen	unpub.
BCD249	barley cDNA	F	M. Sorrells	Heun et al. (1991)
ksuE18	<i>T. tauschii</i> gDNA	F	B.S. Gill	Gill et al. (1991)
ksuG9	"	G	"	"
CDO99	oat cDNA	G	M. Sorrells	Heun et al. (1991)
BCD98	barley cDNA	E	"	"
ksuD14B	<i>T. tauschii</i> gDNA	G	B.S. Gill	Gill et al. (1991)
CDO442	oat cDNA	G	M. Sorrells	Heun et al. (1991)
WG180A	wheat gDNA	G	"	"
WG983	"	F	"	"
BCD304	barley cDNA	G	"	"
BCD265	"	G	"	"
CDO989	oat cDNA	G	"	"
ksuE2B	<i>T. tauschii</i> gDNA	F	B.S. Gill	Gill et al. (1991)
CDO394B	oat cDNA	P	M. Sorrells	Heun et al. (1991)
CDO393	"	E	"	"
ksuE11	<i>T. tauschii</i> gDNA	G	B.S. Gill	Gill et al. (1991)
WG241	wheat gDNA	G	M. Sorrells	Heun et al. (1991)
Vada by <i>Hordeum vulgare</i> subsp. <i>spontaneum</i> line 1B-87 Cross (V/H.s.)				
pBSC4	B-hordeins	?	A. Graner	Graner et al. (1991)
MWG60	barley gDNA	?	"	"
MWG36	"	?	"	"
MWG68	"	?	"	"
pBSC5	C-hordeins	?	"	"
MWG75	barley gDNA	?	"	"

Table 5 Continued

Designation	Description	Rating ^b	Source	Reference
MWG62	„	?	„	„
MWG20	„	?	„	„
MWG55	„	?	„	„
MWG45	„	?	„	„
MWG12	„	?	„	„
MWG56	„	?	„	„
MWG78	„	?	„	„
MWG518	„	?	„	„
MWG504	„	?	„	„
MWG92	„	?	„	„
cMWG706	barley cDNA	?	„	„
MWG824	barley gDNA	?	„	„
cMWG710b	barley cDNA	?	„	„
Igri by Franka Cross (I/F)				
pBSC4	B-hordens	?	A Graner	Graner et al (1991)
cMWG645	barley cDNA	?	„	„
Mla6	powdery mildew resistance	?	„	„
pBSC5	C-hordens	?	„	„
cMWG758	barley cDNA	?	„	„
MWG800	barley gDNA	?	„	„
MWG506	„	?	„	„
cMWG649b	barley cDNA	?	„	„
cMWG676	„	?	„	„
cMWG733	„	?	„	„
cMWG701b	„	?	„	„
MWG632b	barley gDNA	?	„	„

^a Markers are arranged in order from the most distal on the plus arm to the most distal on the minus arm within each population used to construct the maps

^b E – excellent, G – good, F – fair, P – poor

^c NABGMP – North American Barley Genome Mapping Project

Since telosomic addition lines were not available for chromosome 5, it has not been possible to locate the centromere. The isolation and availability of a chromosome 5 plus arm telosomic addition line was recently reported (Shepherd and Islam 1992). This and other cytogenetic stocks may be useful for resolving this issue in the future.

4.6 Chromosome 6 (6H)

A total of 71 markers have been placed on the chromosome 6 RFLP maps (Table 6). Since there are 12 markers identifying probable identical loci on two maps, there are 59 unique markers available for chromosome 6. Most of the markers on the S/M map are also mapped on the S/W map with excellent

CHROMOSOME 6 (6H)

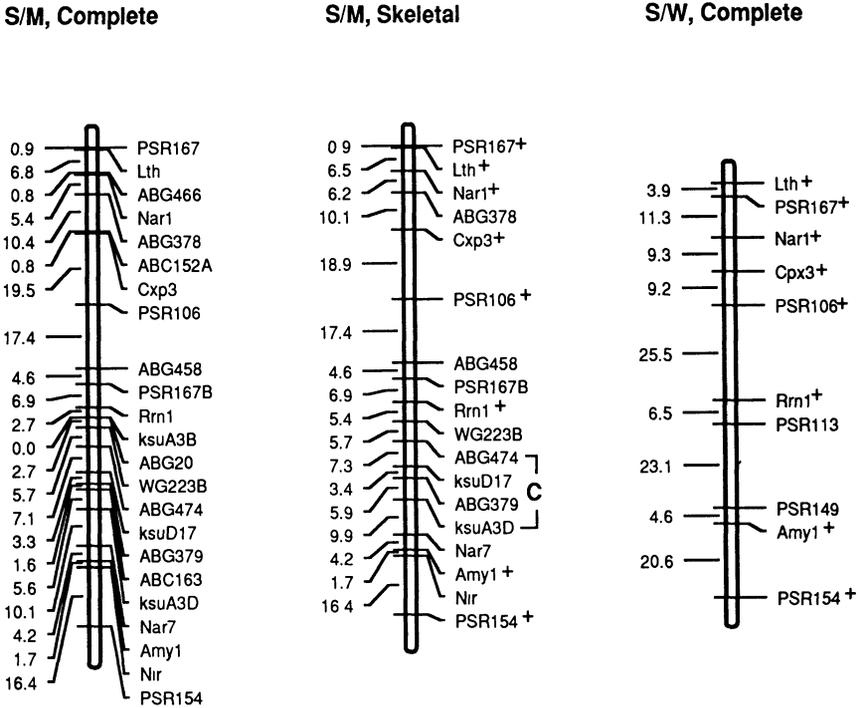


Fig. 4. Chromosome 6 (6H) map, based on the Steptoe by Morex (S/M) cross is compared to the map developed with the Steptoe mutant Az12 by Winer (S/W) cross. Skeletal maps are presented to facilitate comparison of marker order and distances. Markers probably identifying the same loci in the S/M and S/W maps are marked with a '+'. The approximate centromere location is marked with a 'C' and a bracket identifying the two most proximal markers tested on each arm.

alignment (Fig. 4). In fact, the two most distal markers on the two maps are essentially the same, indicating that no progress has been made in extending the genetic length of this map in the last few years. The P/N chromosome 6 map represents only a small fraction of the chromosome and can not be aligned with any of the other maps. The V/H.s. and I/F maps share 4 common markers which nearly define the most distal ends of both maps (Graner et al. 1991). This suggests that the increased length of the V/H.s. map compared to the other RFLP maps may be due to increased recombination in the cultivated by wild barley cross. The S/M and S/W maps have the *Amy1* marker in common with the classical map. This is helpful for orientation of these maps with the classical map and for approximation of the centromere position. The *Amy1* locus has been previously mapped 26 and 23 recombination units (Linde-Laursen 1979, 1982) from a centromeric C-band on 6S. This places the centromere in the vicinity of ABC163 on the S/M map. Based on telosomic addition line data, the centromere

is located between markers ABG474 and ksuA3D on the S/M map (Fig. 4). This centromere location suggests a lack of markers and a very short genetic distance for the minus arm of chromosome 6. Furthermore, since the ribosomal RNA gene is located at the secondary constriction just proximal to the satellite (Appels et al. 1980), these data suggest that recombination between the centromere and the satellite on the plus arm of chromosome 6 is limited, as previously noted (Kleinhofs et al. 1988; Linde-Laursen 1982; Dvorak and Chen 1984; Dvorak and Appels 1986). Chromosome 6 telomeric regions have not yet been mapped. However, the close agreement among the different RFLP maps suggests that either they are nearly full length or the terminal regions of chromosome 6 are very deficient in markers.

Table 6 Markers mapped to barley chromosome 6 (6H) ^a

Designation	Description	Rating ^b	Source	Reference
Stephoe by Morex Cross (S/M)				
PSR167	wheat leaf cDNA	G	M Gale	Chao et al (1988)
Lth	leaf thionins	G	K Apel	Bohlmann and Apel (1987)
ABG466	barley gDNA	G	NABGMP ^c	Kleinhofs (1992)
Nar1	NADH nitrate reductase	E	A Kleinhofs	Schnorr et al (1991)
ABG378	barley gDNA	E	NABGMP	Kleinhofs (1992)
ABC152A	barley malt cDNA	G	"	"
Cxp3	carboxypeptidase	E	D Baulcombe	Baulcombe et al (1987)
PSR106	wheat leaf cDNA	G	M Gale	Chao et al (1988)
ABG458	barley gDNA	G	NABGMP	Kleinhofs (1992)
PSR167B	wheat leaf cDNA	G	M Gale	Chao et al (1988)
Rrn1	ribosomal RNA	E	M O'Dell	Gerlach and Bedbrook (1979)
ABG20	barley gDNA	E	NABGMP	Kleinhofs (1992)
ksuA3B	<i>T. tauschii</i> gDNA	G	B S Gill	Gill et al (1991)
WG223B	wheat gDNA	F	M Sorrells	Heun et al (1991)
ABG474	barley gDNA	E	NABGMP	Kleinhofs (1992)
ksuD17	<i>T. tauschii</i> gDNA	G	B S Gill	Gill et al (1991)
ABG379	barley gDNA	E	NABGMP	Kleinhofs (1992)
ABC163	barley malt cDNA	E	"	"
ksuA3D	<i>T. tauschii</i> gDNA	G	B S Gill	Gill et al (1991)
Nar7	NAD(P)H nitrate reductase	E	A Kleinhofs	Miyazaki et al (1991)
Amy1	α -amylase	G	J Rogers	Khurshheed and Rogers (1988)
Nir	nitrite reductase	G	S Rothstein	Lahners et al (1988)
PSR154	wheat leaf cDNA	G	M Gale	Chao et al (1988)
Proctor by Nudinka Cross (P/N)				
ksuF2	<i>T. tauschii</i> gDNA	F	B S Gill	Gill et al (1991)
BG140A	barley gDNA	F	N Lapitan	Heun et al (1991)
WG282	wheat gDNA	G	M Sorrells	"
BCD269	barley cDNA	E	"	"

Table 6 Continued

Designation	Description	Rating ^b	Source	Reference
CDO419A	oat cDNA	F	"	"
WG222B	wheat gDNA	G	"	"
Dhn3,4	dehydrin	-	T Close	Close and Chandler (1990)
BCD339B	barley cDNA	F	M Sorrells	Heun et al (1991)
BCD221A	"	F	"	"
Vada by <i>Hordeum vulgare</i> subsp. <i>spontaneum</i> line 1B-87 Cross (V/H.s.)				
MWG620	barley gDNA	?	A Graner	Graner et al (1991)
MWG59	"	?	"	"
MWG10a	"	?	"	"
cMWG652	barley cDNA	?	"	"
cMWG690	"	?	"	"
cMWG653	"	?	"	"
cMWG664	"	?	"	"
MWG549	barley gDNA	?	"	"
MWG26	"	?	"	"
MWG79b	"	?	"	"
cMWG664	barley cDNA	?	"	"
cMWG666	"	?	"	"
cMWG679	"	?	"	"
MWG19	barley gDNA	?	"	"
cMWG669	barley cDNA	?	"	"
MWG38	barley gDNA	?	"	"
cMWG674	barley cDNA	?	"	"
MWG21	barley gDNA	?	"	"
MWG74	"	?	"	"
MWG514	"	?	"	"
Igri by Franka Cross (I/F)				
MWG573	barley gDNA	?	A Graner	Graner et al (1991)
cMWG652	barley cDNA	?	"	"
cMWG679	"	?	"	"
MWG820	barley gDNA	?	"	"
cMWG669	barley cDNA	?	"	"
cMWG684c	"	?	"	"
cMWG716b	"	?	"	"
MWG549a	barley gDNA	?	"	"
cMWG684a	barley cDNA	?	"	"
MWG514	barley gDNA	?	"	"
Steptoe mutant Az12 by Winer Cross (S/W)				
PSR167	wheat leaf cDNA	G	M Gale	Kleinhofs et al (1988)
Nar1	NADH nitrate reductase	E	A Kleinhofs	Schnorr et al (1991)
Cxp3	carboxypeptidase	E	D Boulcombe	Boulcombe et al (1987)
PSR106	wheat leaf cDNA	G	M Gale	Kleinhofs et al (1988)
Rrn1	ribosomal RNA	E	M O'Dell	Gerlach and Bedbrook (1979)

Table 6 Continued

Designation	Description	Rating ^b	Source	Reference
PSR113	wheat leaf cDNA	G	M Gale	Kleinhofs et al (1988)
PSR149	"	G	"	"
Amy1	α -amylase	G	D C Baulcombe	Lazarus et al (1985)
PSR154	wheat leaf cDNA	G	M Gale	Kleinhofs et al (1988)

^a Markers are arranged in order from the most distal on the plus arm to the most distal on the minus arm within each population used to construct the maps

^b E – excellent, G – good, F – fair, P – poor

^c NABGMP – North American Barley Genome Mapping Project

4.7. Chromosome 7 (5H)

A total of 111 markers have been placed on the chromosome 7 RFLP maps (Table 7). Since 15 markers identify probable identical loci on two maps, there are 96 unique markers available for chromosome 7. The S/M and P/N maps share 10 markers, allowing reasonable alignment of the two maps (Fig. 5). The V/H.s. and I/F maps have 5 common markers and have been previously aligned (Graner et al. 1991). The molecular maps do not share any markers with the classical map. The chromosome 7 centromere is located between the *Rrn2* and *Ipa* loci (Fig. 5). Thus, chromosome 7 appears to have a genetically long and well marked minus arm while the plus arm appears to be deficient in markers. Chromosome 7 telomeric sequences have not yet been mapped

Table 7 Markers mapped to barley chromosome 7 (5H) ^a

Designation	Description	Rating ^b	Source	Reference
Stephoe by Morex Cross (S/M)				
ABC483	barley malt cDNA	G	NABGMP ^c	Kleinhofs (1992)
ABR313	CG TA CG CG TT (primer)	F	"	"
Adh6	alcohol deh , cDNA	P	A Good	Good et al (1988)
CDO749	oat cDNA	E	M Sorrells	Heun et al (1991)
Rrn 2	ribosomal RNA	F	M O'Dell	Gerlach and Bedbrook (1979)
Ipa	amylase/protease inhibitor	G	J Rogers	Mundy and Rogers (1986)
ksuA3A	<i>T. tauschii</i> gDNA	G	B S Gill	Gill et al (1991)
WG541	wheat gDNA	G	M Sorrells	Heun et al (1991)
WG530	"	G	"	"
BG123B	barley gDNA	F	N Lapitan	"
WG889	wheat gDNA	F	M Sorrells	"
Ale	aleurain	E	J Rogers	Rogers et al (1985)
CDO348B	oat cDNA	G	M Sorrells	Heun et al (1991)
PSR128	wheat cDNA	F	M Gale	Chao et al (1988)
ABC302	barley malt cDNA	E	NABGMP	Kleinhofs (1992)

Table 7. Continued.

Designation	Description	Rating ^b	Source	Reference
CDO57B	oat cDNA	G	M. Sorrells	Heun et al. (1991)
ABG473	barley gDNA	F	NABGMP	Kleinhofs (1992)
WG364	wheat gDNA	G	M. Sorrells	Heun et al. (1991)
BCD351E	barley cDNA	F	"	"
ksuA1B	<i>T. tauschii</i> gDNA	G	B.S. Gill	Gill et al. (1991)
mSrh	rachila hairs, morphological	E	J. Franckowiak	Haus (1978)
CDO504	oat cDNA	G	M. Sorrells	Heun et al. (1991)
WG644	wheat gDNA	F	"	"
iEst9	esterase, isozyme	E	P. Hayes	Nielsen and Johansen (1986)
WG908	wheat gDNA	G	M. Sorrells	Heun et al. (1991)
ABG495A	barley gDNA	G	NABGMP	Kleinhofs (1992)
ABC155	barley malt cDNA	E	"	"
Cab1	chlorophyll a/b binding	F	K. Gausing	Barkardottir et al. (1987)
ABC482	barley cDNA	G	W. Rohde	unpub.)
ABC310A	barley malt cDNA	F	NABGMP	Kleinhofs (1992)
CDO484	oat cDNA	E	M. Sorrells	Heun et al. (1991)
ABG463	barley gDNA	E	NABGMP	Kleinhofs (1992)
ABC309	barley malt cDNA	G	"	"
Proctor by Nudinka Cross (P/N)				
BG140C	barley gDNA	F	N. Lapitan	Heun et al. (1991)
CDO749	oat cDNA	G	M. Sorrells	"
rDNA	ribosomal RNA	F	?	Gerlach and Bedbrook (1979)
CDO460	oat cDNA	P	M. Sorrells	Heun et al. (1991)
WG541	wheat gDNA	G	"	"
WG530	"	E	"	"
WG889	"	G	"	"
WG564	"	E	"	"
WG364	"	F	"	"
ksuE2A	<i>T. tauschii</i> gDNA	F	B.S. Gill	Gill et al. (1991)
CDO675B	oat cDNA	F	M. Sorrells	Heun et al. (1991)
ksuA3	<i>T. tauschii</i> gDNA	G	B.S. Gill	Gill et al. (1991)
CDO771A	oat cDNA	F	M. Sorrells	Heun et al. (1991)
WG1026	wheat gDNA	G	"	"
CDO504	oat cDNA	G	"	"
WG644	wheat gDNA	F	"	"
CDO583	oat cDNA	F	"	"
WG908A	wheat gDNA	G	"	"
WG908B	"	G	"	"
BCD298	barley leaf cDNA	F	"	"
CDO113	oat cDNA	F	"	"
CDO213	"	G	"	"
CDO457	"	P	"	"
CDO484	"	E	"	"
CDO506	"	E	"	"

Table 7. Continued.

Designation	Description	Rating ^b	Source	Reference
Vada by <i>Hordeum vulgare</i> subsp. <i>spontaneum</i> line 1B-87 Cross (V/H.s.)				
MWG15	barley gDNA	?	A. Graner	Graner et al. (1991)
MWG502	"	?	"	"
MWG63	"	?	"	"
MWG28	"	?	"	"
cMWG718a	barley leaf cDNA	?	"	"
MWG25	barley gDNA	?	"	"
MWG596	"	?	"	"
MWG526	"	?	"	"
MWG609	"	?	"	"
MWG561b	"	?	"	"
MWG16	"	?	"	"
MWG40	"	?	"	"
MWG998	"	?	"	"
MWG37	"	?	"	"
MWG13	"	?	"	"
MWG7	"	?	"	"
MWG592	"	?	"	"
MWG522	"	?	"	"
MWG624	"	?	"	"
MWG50	"	?	"	"
MWG805	"	?	"	"
MWG72	"	?	"	"
MWG516	"	?	"	"
MWG76	"	?	"	"
MWG79a	"	?	"	"
MWG34	"	?	"	"
MWG2	"	?	"	"
MWG813	"	?	"	"
Igri by Franka Cross (I/F)				
MWG502	barley gDNA	?	A. Graner	Graner et al. (1991)
MWG618	"	?	"	"
cMWG717	barley leaf cDNA	?	"	"
MWG526	barley gDNA	?	"	"
MWG561b	"	?	"	"
MWG596	"	?	"	"
cMWG770	barley leaf cDNA	?	"	"
MWG522	barley gDNA	?	"	"
MWG583	"	?	"	"
MWG624	"	?	"	"
MWG999	"	?	"	"
MWG604	"	?	"	"
MWG549b	"	?	"	"
MWG553a	"	?	"	"
MWG550	"	?	"	"
cMWG646b	barley leaf cDNA	?	"	"
MWG533	barley gDNA	?	"	"
cMWG716a	barley leaf cDNA	?	"	"

Table 7. Continued.

Designation	Description	Rating ^b	Source	Reference
MWG827	barley gDNA	?	„	„
cMWG701a	barley leaf cDNA	?	„	„
cMWG654	„	?	„	„
cMWG740	„	?	„	„
cMWG650	„	?	„	„
MWG602	barley gDNA	?	„	„
MWG813	„	?	„	„

^a Markers are arranged in order from the most distal on the plus arm to the most distal on the minus arm within each population used to construct the maps.

^b E – excellent; G – good; F – fair; P – poor.

^c NABGMP – North American Barley Genome Mapping Project.

5. Anomalous probes

Anomalous mapping behavior was observed with probe BCD351. This probe revealed a high level of polymorphism and mapped to numerous loci among barley cultivars. The four cultivars used by NABGMP (Steptoe, Morex, Harrington, TR306) had unique banding patterns for 6 restriction enzymes tested. There was also consistent variation among these genotypes in the number of bands obtained with various enzymes. These results suggested that there is a difference in the number of copies of the sequence homologous to the probe among barley cultivars. The segregation pattern obtained while mapping the DH lines from the S/M cross showed that the 3 major bands were not allelic and mapped to 3 different chromosomes. An additional faint band was also mapped to a separate locus.

The loci detected by the BCD351 probe and mapped with the S/M cross were compared to those mapped with the P/N cross. Two loci mapped with the P/N cross reside on chromosome 1 and chromosome 2 (Heun et al. 1991). Neither appears to be the same location as the four loci mapped on the S/M cross. Even the locus found on chromosome 2 on both maps appears to be in a different position. This locus, designated BCD351F, is located on the S/M map between markers WG516 and CDO64 while locus BCD351B is proximal to CDO64 on the P/N map. Both the order and genetic distances between WG516 and CDO64 are conserved on both maps (Fig. 2).

Sequences homologous to BCD351 were mapped to chromosomal arms of cv. Betzes by hybridization to restriction digests of ditelosomic addition lines (Islam 1983) (Betzes is the source of barley chromosomes added to the genetic background of wheat in these lines). The pattern of hybridization indicated that BCD351 homologous sequences are found on chromosomes 1M, 4P, 7M, and probably 5 in cv. Betzes compared to chromosomes 2P, 4P, 5, and 7M in the cvs. Steptoe and Morex. These results indicate that the sequences homologous to

CHROMOSOME 7 (5H)

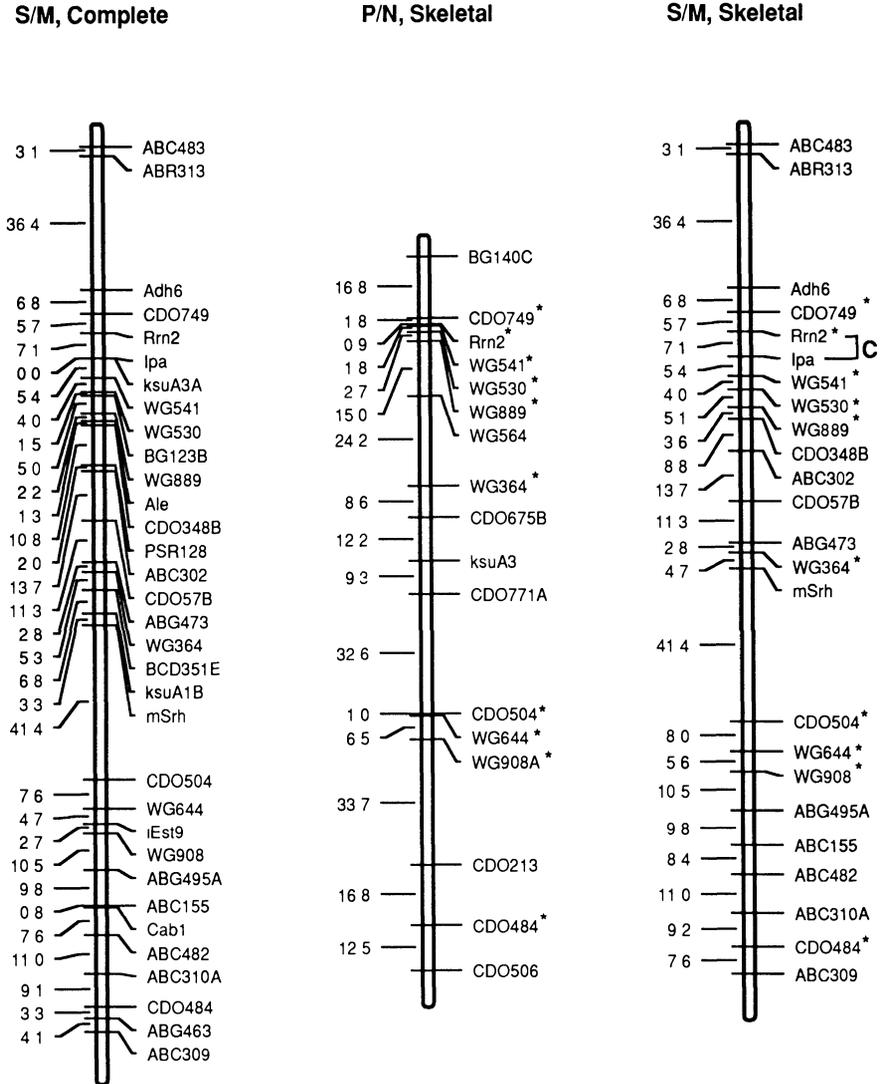


Fig 5 Chromosome 7 (5H) map, based on the Steptoe by Morex (S/M) cross is compared to the map developed with the Proctor by Nudinka (P/N) cross. Skeletal maps are presented to facilitate comparison of marker order and distances. Markers probably identifying the same loci in the S/M and P/N maps are marked with a '*'. The approximate centromere location is marked with a 'C' and a bracket identifying the two most proximal markers tested on each arm.

Table 8. Summary of the number of markers (M) and cM mapped for each chromosome (Chr) with several different crosses.

Chr	Cross ^a									
	S/M		P/N		V/H.s.		I/F		S/W	
	M	cM	M	cM	M	cM	M	cM	M	cM
1	40	230	32	189	19	218	26	186	10	160
2	36	231	38	194	29	230	22	225	–	–
3	24	197	14	195	30	223	18	207	–	–
4	18	159	16	114	18	180	4	32	–	–
5	26	176	20	163	19	152	13	109	–	–
6	23	134	9	39	20	174	10	135	9	110
7	33	251	25	197	28	226	25	245	–	–
Total	200	1,378	154	1,091	163	1,403	118	1,139	19	–

^a The crosses are S/M – Steptoe by Morex; P/N – Proctor by Nudinka; V/H.s. – Vada by *Hordeum vulgare* subsp. *spontaneum*; I/F – Igrı by Franka; S/W – Steptoe mutant Az12 by Winer.

BCD351 underwent frequent rearrangements in the barley genome. These rearrangements were relatively recent, as indicated by the differences in the number of copies and the location of these sequences among barley cultivars and are suggestive of a transposable element. The mechanism responsible for multiplication and spread of sequences homologous to BCD351 within the barley genome remains to be determined.

6. Summary

Barley RFLP map development has been explosive during the past few years. As of this writing, there are nearly 600 RFLP markers mapped to the seven barley chromosomes and the number is increasing daily (Table 8). There are doubled haploid, therefore ‘immortal’, populations available for future mapping. The NABGMP, in particular, has developed several very useful populations that should serve the barley genetics community for many years to come. Additional work is needed to identify all telomeric regions, merge the different RFLP maps and correlate the RFLP maps with classical and physical maps. The current and future development of barley maps should facilitate the application of genetics to barley breeding and the use of barley as a genetic organism for further basic studies.

7. Acknowledgements

Research in A. Kleinhofs' laboratory is supported in part by CSRS Special Grant Agreement No. 90-34213-5190, Washington Technology Center and

Washington Barley Commission. This publication is Scientific Paper Number 9201-32, College of Agriculture and Home Economics Research Center, Washington State University, Pullman, WA, U.S.A. Project NO. 0951.

8. References

- Appels, R, Gerlach, W L, Dennis, E S, Swift, H and Peacock, W J (1980) Molecular and chromosomal organization of DNA sequences coding for the ribosomal RNAs in cereals *Chromosoma* 78 293-311
- Asher, M J C, Ellis, R P, Hayter, A M and Whitehouse, R N H (eds) (1981) *Barley Genetics IV* Edinburgh University Press, Edinburgh
- Barkardottir, R B, Jensen, B F, Kreiberg, J D, Nielsen, P S and Gausing, K (1987) Expression of selected nuclear genes during leaf development in barley *Dev Genet* 8 495-511
- Baulcombe, D C, Barker, R F and Jarvis, M G (1987) A gibberellin responsive wheat gene has homology to yeast carboxypeptidase Y *J Biol Chem* 262 13726-13735
- Bennett, M D and Smith, J B (1976) Nuclear DNA amounts in angiosperms *Phil Trans R Soc London (Biol)* 274 227-274
- Bohlmann, H and Apel, K (1987) Isolation and characterization of cDNAs coding for leaf-specific thionins closely related to the endosperm-specific hordothionin of barley (*Hordeum vulgare* L) *Mol Gen Genet* 207 446-454
- Briggs, D E (1978) *Barley* Chapman & Hall, London
- Burnham, C R and Hagberg, A (1956) Cytogenetic notes on chromosomal interchanges in barley *Hereditas* 42 467-482
- Cameron-Mills, V and Brandt, A (1988) A γ -hordein gene *Plant Mol Biol* 11 449-461
- Chao, S, Sharp, P J and Gale, M D (1988) A linkage map of wheat homoeologous group 7 chromosomes using RFLP markers In T E Miller and R M D Koebner (eds), *Proc 7th Int Wheat Genet Symp*, pp 493-498 Bath Press, Bath
- Chao, S, Raines, C A, Longstaff, M, Sharp, P J, Gale, M D and Dyer, T A (1989) Chromosomal location and copy number in wheat and some of its close relatives of genes for enzymes involved in photosynthesis *Mol Gen Genet* 218 423-430
- Chen, F and Hayes, P M (1989) A comparison of *Hordeum bulbosum*-mediated haploid production efficiency in barley using in vitro floret and tiller culture *Theor Appl Genet* 77 701-704
- Chojacki, J (1986) Identification and characterization of a cDNA clone for histone H3 in barley *Carlsberg Res Commun* 51 211-217
- Clark, H H (1967) The origin and early history of the cultivated barleys *Agric Hist Rev* 15 1-18
- Close, T J and Chandler, P M (1990) Cereal dehydrins serology, gene mapping and potential functional roles *Aust J Plant Physiol* 17 333-344
- Dvorak, J and Appels, R (1986) Investigation of homologous crossing over and sister chromatid exchange in the wheat *Nor-B2* locus coding for rRNA and *Gli-B2* locus coding for gliadins *Genetics* 113 1037-1056
- Dvorak, J and Chen, K -C (1984) Distribution of nonstructural variation between wheat cultivars along chromosome arm 6Bp evidence from the linkage map and physical map of the arm *Genetics* 106 325-333
- Forde, B G, Kreis, M, Williamson, M S, Fry, R P, Pywell, J, Shewry, P R, Bunce, N and Miflin, B J (1985) Short tandem repeats shared by B- and C-hordein cDNAs suggest a common evolutionary origin for two groups of cereal storage protein genes *EMBO J* 4 9-15
- Freeman, J, Marquez, A, Wallsgrove, R M, Saarelainen, R and Forde, B G (1990) Molecular analysis of barley mutants deficient in chloroplast glutamine synthetase *Plant Mol Biol* 14 297-311
- Gaul, H (ed) (1976) *Barley Genetics III* Verlag Karl Thiemig, Munich
- Gausing, K and Barkardottir, R (1986) Structure and expression of ubiquitin genes in higher plants *Eur J Biochem* 158 57-62

- Gerlach, W L and Bedbrook, J R (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley Nucl Acids Res 7 1869-1885
- Gill, K S , Lubbers, E L , Gill, B S , Raupp, W J and Cox, T S (1991) A genetic linkage map of *Triticum tauschii* (DD) and its relationship to the D genome of bread wheat (AABBDD) Genome 34 362-374
- Good, A G , Pelcher, L E and Crosby, W L (1988) Nucleotide sequence of a complete barley alcohol dehydrogenase 1 cDNA Nucl Acids Res 16 7182
- Graner, A , Jahoor, A , Schondelmaier, J , Siedler, H , Pillen, K , Fischbeck, G , Wenzel, G and Herrmann, R G (1991) Construction of an RFLP map of barley Theor Appl Genet 83 250-256
- Harlan, J R (1976) Barley, *Hordeum vulgare* (Graminea-triticeae) In N W Simmonds (ed), Evolution of Crop Plants, pp 93-98 Longman Group, London
- Harlan, J R (1979) On the origin of barley In Barley Origin, Botany, Culture, Winter Hardiness, Genetics, Utilization, Pests, pp 10-36 USDA Agric Handbook 338
- Haus, T E (1978) Short-haired rachilla Barley Genet Newsletter 8 160
- Hertig, C , Rebmann, G , Bull, J , Mauch, F and Dudler, R (1991) Sequence and tissue-specific expression of a putative peroxidase gene from wheat (*Triticum aestivum* L) Plant Mol Biol 16 171-174
- Heun, M , Kennedy, A E , Anderson, J A , Lapitan, N L V , Sorrells, M E and Tanksley, S D (1991) Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*) Genome 34 437-447
- Hinze, K , Thopson, R D , Ritter, E , Salamini, F and Schulze-Lefert, P (1991) Restriction fragment length polymorphism-mediated targeting of the *ml-o* resistance locus in barley (*Hordeum vulgare*) Proc Natl Acad Sci U S A 88 3691-3695
- Islam, A K M R (1980) Identification of wheat-barley addition lines with N-banding of chromosomes Chromosoma 76 365-373
- Islam, A K M R (1983) Ditelosomic additions of barley chromosomes to wheat In S Sakamoto (ed), Proc 6th Int Wheat Genet Symp , pp 233-238 Kyoto University Press, Kyoto
- Kahler, A L and Allard, R W (1970) Genetics of isozyme variants in barley 1 Esterases Crop Sci 10 444-448
- Kasha, K J (1974) Haploids from somatic cells In K J Kasha (ed), Haploids in Higher Plants Advances and Potential, pp 67-87 University of Guelph, Guelph
- Kasha, K J and Kao, K N (1970) High frequency haploid production in barley Nature 225 874-876
- Khursheed, B and Rogers, J C (1988) Barley α -amylase genes Quantitative comparison of steady-state mRNA levels from individual members of the two different families expressed in aleurone cells J Biol Chem 263 18953-18960
- Kleinhofs, A (1992) The NABGMP mapping progress report Barley Genet Newsl 21 38-48
- Kleinhofs, A , Chao, S and Sharp, P J (1988) Mapping of nitrate reductase genes in barley and wheat In T E Miller and R M D Koebner (eds), Proc 7th Int Wheat Genet Symp , pp 541-546 Bath Press, Bath
- Kreis, M , Williamson, M , Buxton, B , Pywell, J , Hejgaard, J and Svendsen, I (1987) Primary structure and differential expression of β -amylase in normal and mutant barleys Eur J Biochem 169 517-525
- Lahners, K , Kramer, V , Back, E , Privalle, L and Rothstein, S (1988) Molecular cloning of complementary DNA encoding maize nitrite reductase Plant Physiol 88 741-746
- Lamberts, H , Broekhuizen, S , Dantuma, G and Lange, W (eds) (1964) Barley Genetics I Pudoc, Wageningen
- Lazarus, C M , Baulcombe, D C and Martienssen, R A (1985) α -Amylase genes of wheat are two multigene families which are differentially expressed Plant Mol Biol 5 13-24
- Leah, R , Tommerup, H , Svendsen, I and Mundy, J (1991) Biochemical and molecular characterization of three barley seed proteins with antifungal properties J Biol Chem 266 1564-1573

- Linde-Laursen, I (1979) Giemsa C-banding of barley chromosomes III Segregation and linkage of C-bands on chromosomes 3, 6, and 7 *Hereditas* 91 73–77
- Linde-Laursen, I (1982) Linkage map of the long arm of barley chromosome 3 using C-bands and marker genes *Heredity* 49 27–35
- Linde-Laursen, I and Jensen, J (1992) Proposal for new designations of chromosome arms and positions of chromosomal markers In L Munck (ed), *Barley Genetics VI*, Vol 1, pp 287–289 Munksgaard Int Pub Ltd , Copenhagen
- Litts, J C , Simmons, C R , Karrer, E E , Huang, N and Rodriguez, R J (1990) The isolation and characterization of a barley 1,3-1,4- β -glucanase gene *Eur J Biochem* 194 831–838
- Miyazaki, J , Juricek, M , Angelis, K , Schnorr, K M , Kleinhofs, A and Warner, R L (1991) Characterization and sequence of a novel nitrate reductase from barley *Mol Gen Genet* 228 329–334
- Morris, C F , Anderberg, R J , Goldmark, P J and Walker-Simmons, M K (1991) Molecular cloning and expression of abscisic acid-responsive genes in embryos of dormant wheat seeds *Plant Physiol* 95 814–821
- Muir, C E and Nilan, R A (1973) Registration of Steptoe barley *Crop Sci* 13 770
- Munck, L (ed) (1992) *Barley Genetics VI* Munksgaard Int Pub Ltd , Copenhagen
- Mundy, J and Rogers, J C (1986) Selective expression of a probable amylase/protease inhibitor in barley aleurone cells Comparison to the barley amylase/subtilisin inhibitor *Planta* 169 51–63
- Nielsen, G and Johansen, H B (1986) Proposal for the identification of barley varieties based on genotypes for 2 hordein and 39 isoenzyme loci of 47 reference varieties *Euphytica* 35 717–728
- Nielsen, P S and Gausung, K (1987) The precursor of barley plastocyanin Sequence of cDNA clones and gene expression in different tissues *FEBS Lett* 225 159–162
- Nilan, R A (1964) *The Cytology and Genetics of Barley, 1951–1962* Monogr Suppl 3, Res Stud Vol 32, No 1, Washington State University Press, Pullman
- Nilan, R A (1974) *Barley (Hordeum vulgare)* In R C King (ed), *Handbook of Genetics*, pp 93–110 Plenum Press, New York
- Nilan, R A (ed) (1971) *Barley Genetics II* Washington State University Press, Pullman
- Olive, M R , Ellis, R J and Schuch, W W (1989) Isolation and nucleotide sequences of cDNA clones encoding ADP-glucose pyrophosphorylase polypeptides from wheat leaf and endosperm *Plant Mol Biol* 12 525–538
- Ramage, R T (1985) *Cytogenetics* In D C Rasmusson (ed), *Barley* pp 127–154 American Society of Agronomy, Publishers, Madison
- Rasmusson, S K, Welinder, K G and Hejgaard, J (1991) cDNA cloning, characterization and expression of an endosperm-specific barley peroxidase *Plant Mol Biol* 16 317–327
- Rasmusson, D C (ed) (1985) *Barley* American Society of Agronomy, Publishers, Madison
- Rasmusson, D C and Wilcoxson, R D (1979) Registration of Morex barley *Crop Sci* 19 293
- Rebmann, G , Mauch, F and Dudler, R (1991) Sequence of a wheat cDNA encoding a pathogen-induced thaumatin-like protein *Plant Mol Biol* 17 283–285
- Rogers, J C , Dean, D and Heck, G R (1985) Aleurain A barley thiol protease closely related to mammalian cathepsin H *Proc Natl Acad Sci U S A* 82 6512–6516
- Rohde, W , Becker, D and Salamini, F (1988) Structural analysis of the waxy locus from *Hordeum vulgare* *Nucl Acids Res* 16 7185–7186
- Rohde, W , Dorr, S , Salamini, F and Becker, D (1991) Structure of a chalcone synthase gene from *Hordeum vulgare* *Plant Mol Biol* 16 1103–1106
- Sarvella, P , Holmgren, B J and Nilan, R A (1958) Analysis of barley pachytene chromosomes *Nucleus* 1 183–204
- Schnorr, K M , Juricek, M , Huang, C , Culley, D and Kleinhofs, A (1991) Analysis of barley nitrate reductase cDNA and genomic clones *Mol Gen Genet* 227 411–416
- Schulz, R , Steinmuller, K , Klaas, M , Forreiter, C , Rasmusson, S , Hiller, C and Apel, K (1989) Nucleotide sequence of a cDNA coding for the NADPH-protochlorophyllide oxidoreductase (PCR) of barley (*Hordeum vulgare* L.) and its expression in *Escherichia coli* *Mol Gen Genet* 217 355–361
- Sharp, P J , Desai, S , Chao, S and Gale, M D (1988) Isolation, characterization and applications

- of a set of 14 RFLP probes identifying each homoeologous chromosome arm in the Triticeae In T E Miller and R M D Koebner (eds), Proc 7th Int Wheat Genet Symp , pp 639–646 Bath Press, Bath
- Shepherd, K W and Islam, A K M R (1992) Progress in the production of wheat-barley addition and recombination lines and their use in mapping the barley genome In P R Shewry (ed), Barley Genetics, Biochemistry, Molecular Biology and Biotechnology, pp 99–114 CAB International, Wallingford
- Shewry, P R (ed) (1992) Barley Genetics, Biochemistry, Molecular Biology and Biotechnology CAB International, Wallingford
- Shin JS, Corpuz L, Chao S and Blake TK (1990) A partial map of the barley genome Genome 33 803–808
- Singh, R J and Tsuchiya, T (1975) Pachytene chromosomes of barley J Hered 66 165–167
- Singh, R J and Tsuchiya, T (1982) Identification and designation of telocentric chromosomes in barley by means of Giemsa N-banding technique Theor Appl Genet 64 13–24
- Smith, L (1951) Cytology and genetics of barley Bot Rev 17 1–51, 133–202, 285–355
- Sogaard, B and von Wettstein-Knowles, P (1987) Barley genes and chromosomes Carlsberg Res Commun 52 123–196
- Tabata, T , Sasaki, K and Iwabuchi, M (1983) The structural organization and DNA sequence of a wheat histone H4 gene Nucl Acids Res 11 5865–5875
- Tsuchiya, T (1960) Cytogenetic studies of trisomics in barley Jap J Bot 17 177–213
- Von Wettstein-Knowles, P (1992) Cloned and mapped genes current status In P R Shewry (ed), Barley Genetics, Biochemistry, Molecular Biology and Biotechnology, pp 73–98 CAB International, Wallingford
- Wendorf, F , Schild, R , El Hadidi, N , Close, A E , Kobusiewicz M , Wieckowska, H , Issawi, B and Haas, H (1979) Use of barley in the Egyptian late Paleolithic Science 205 1341–1347
- Williamson, M S , Forde, J and Kreis, M (1988) Molecular cloning of two isoforms of chymotrypsin inhibitor 1 (CI-1) from barley endosperm and their expression in normal and mutant barleys Plant Mol Biol 10 521–535
- Wise, R P , Rohde, W and Salamini, F (1990) Nucleotide sequence of the *Bronze-1* homologous gene from *Hordeum vulgare* Plant Mol Biol 14 277–279
- Yasuda, S and Konishi, T (eds) (1987) Barley Genetics V Sanyo Press, Okayama

11. DNA-based marker maps of *Brassica*

CARLOS F. QUIROS, JINGUO HU and MARIA JOSE TRUCO¹

Department of Vegetable Crops, University of California, Davis, CA 95616 U S A ¹ *Present address IRTA, Crta de Cabrils s/n, 08348 Cabrils, Spain*

Contents

1	Introduction	199	3 1 2	B genome maps	208
2	Mapping tools in <i>Brassica</i>	201	3 2	F ₂ linkage maps	208
2 1	Cytogenetic stocks	201	3 2 1	<i>B oleracea</i> maps	208
2 2	Genetic markers	201	3 2 2	<i>B rapa</i> maps	209
2 2 1	RFLP markers	201	3 2 3	<i>B nigra</i> map	209
2 2 1 1	Anonymous genomic clones	201	3 2 4	<i>B napus</i> map	213
2 2 1 2	cDNA clones	202	4	Structure of the <i>Brassica</i> genomes	213
2 2 1 3	Gene specific probes	202	4 1	Genomic composition	213
2 2 1 4	Degree of polymorphism	203	4 2	On the origin of the genomes	215
2 2 2	RAPD markers	204	4 3	Plasticity of the <i>Brassica</i> genomes	216
2 2 2 1	Degree of polymorphism	204	4 3 1	Intragenomic homoeologous recombination	216
3	Maps	205	4 3 2	Intergenomic homoeologous recombination	217
3 1	Synteny maps	205	5	Applications of the maps in breeding	218
3 1 1	C genome maps	205	6	References	218
3 1 1 1	Determination of general stability of C genome alien chromosomes	206			

1. Introduction

Brassica crops encompass many diverse types of plants, grown as vegetables, fodder or sources of oils and condiments. The utilization of oilseed *Brassic*as is steadily increasing. At the present time it represents 13.2% of the world's edible oil, with a production of 210 million metric tons (Carr and McDonald 1991). Vegetable *Brassic*as also have great economic importance because of their wide popularity. For example, in California alone they have an annual value of \$460

million. The development of basic genetic information and its application to breeding is appropriate due to the increasing importance of these crops.

The diploid species of *Brassica* range in genomic numbers from $n = 7$ to $n = 12$. The three diploid cultivated species are:

1. *B. nigra* ($2n = 2x = 16$, B genome) black mustard,
2. *B. oleracea* ($2n = 2x = 18$, C genome) cabbage group and
3. *B. rapa* (syn. *campestris*, $2n = 2x = 20$, A genome) turnip, rapeseed and oriental vegetables.

Amphidiploidy and aneuploidy have played important roles during the differentiation and evolution of *Brassica* species (Prakash and Hinata 1980). The genome relationships among cultivated diploid and derived amphidiploid species were elucidated by U (1935). The three basic diploid cultivated species mentioned above have originated the three amphidiploids, *B. carinata* ($2n = 4x = 34$, genomes BC), *B. juncea* ($2n = 4x = 36$, genomes AB) and *B. napus* ($2n = 4x = n = 38$, genomes AC).

Relatedness between the three basic *Brassica* genomes, A, B and C, has been cytogenetically investigated by using digenomic diploids from interspecific hybrids (Attia and Robbelen 1986), from amphidiploids through haploidy (Morinaga and Fukushima 1933; Olsson and Hagberg 1955), and digenomic triploids from crosses between amphidiploids and diploids (Morinaga 1929, 1934; Attia et al. 1987). In general, the three genomes are considered to be mutually and partially homologous and presumably derived from a common ancestral genome (Mizushima 1950). The advent of modern molecular techniques is playing an important role in understanding the organization and relationships of the *Brassica* genomes. These include isozyme (Vaughan 1977; Arus and Orton 1983; Coulthart and Denford 1982; Quiros et al. 1985), chloroplast DNA (Erickson et al. 1983, Palmer et al. 1983; Warwick and Black 1991), nuclear DNA RFLP (Hosaka et al. 1990; Quiros et al. 1987; Song et al. 1988a) and RAPD (Quiros et al. 1991) studies. Results from the above analyses not only confirmed the origin of the amphidiploids but also suggested that the A and C genomes are closely related to each other forming a single lineage, whereas the B genome is genetically distant to both A and C genomes forming a separate lineage (Song et al. 1990; Warwick and Black 1991).

A common assumption is that the $n = 8$, 9 and 10 cultivated species have evolved in an ascending diploid series from a common primitive genome, 'Urgenome' (Haga 1938). Although there are no known *Brassica* species in nature with genomes of $n = 6$, Sikka (1940) and Robbelen (1960) postulated that the ancestral genome consisted of 6 basic chromosomes, which through polysomy, originated the $n = 8$, 9 and 10 chromosome genomes. Thus, as a corollary of this hypothesis, the cultivated diploids are considered secondary polyploids (Prakash and Hinata 1980).

The development of genetic maps in *Brassica* will serve a double purpose: 1) understanding the relationship among the genomes of the *Brassica* diploid cultivated species, and 2) utilization in applied genetics and breeding of the *Brassica* crops.

2. Mapping tools in *Brassica*

2.1. Cytogenetic stocks

In gene mapping projects, cytogenetic stocks provide the means to assign genes and linkage groups to specific chromosomes. Although early research in *Brassica* has involved extensively interploidy crossing, few attempts have been made to develop cytogenetic stocks from the resulting aneuploid progenies. One of these attempts was the creation of three disomic addition lines of *B. rapa* by Fantes and Mackay (1979). Fan et al. (1985) developed a monosomic addition of *B. napus* carrying an alien chromosome from *Diplotaxis muralis*. The first systematic attempt to develop a complete series of alien addition lines for the different *Brassica* genomes was that of Quiros et al. (1986, 1987). Later, others generated alien addition lines for various genomes (Kaneko et al. 1987; Jahier et al. 1989; Chen et al. 1990; Struss et al. 1991a). Several monosomics were generated and studied by Fan and Tai (1985) and Chang et al. (1987) in *B. napus*. Banga (1988) found spontaneous C genome chromosome substitution lines in *B. juncea* (genome AABB) in advanced hybrid generations. Putative C genome substitutions in *B. rapa* were also observed by McGrath et al. (1990) in a few progenies derived from alien addition lines.

2.2. Genetic markers

Although morphological markers have been used for genetic analysis in some of the *Brassica* species (Sampson 1966), they have had minimal impact in gene mapping because of their small numbers. The study of Arus and Orton (1983) on the inheritance and linkage of isozyme loci in *B. oleracea* represented the first major attempt to develop genetic markers and use them to assemble linkage maps in *Brassica*. Later, the advent of DNA based genetic markers such as RFLPs (Figdore et al. 1988; Hosaka et al. 1990; Slocum et al. 1990; Landry et al. 1991) and RAPDs (Quiros et al. 1991) provided sufficient markers to develop comprehensive maps for the *Brassica* genomes.

In the present chapter we will consider only DNA based markers, namely RFLPs and RAPDs.

2.2.1. RFLP markers

In *Brassica* RFLP markers have been derived from anonymous genomic and cDNA clones and gene specific sequences.

A general description on the origin of these markers and their use in mapping follows.

2.2.1.1. *Anonymous genomic clones.* Most of the anonymous genomic clones used in *Brassica*, approximately 370, were derived from several *Pst*I libraries (Figdore et al. 1988; Song et al. 1988a,b). These clones range from 500 to 2000 bp and contain low copy number DNA sequences. They were isolated from

B. oleracea 'Wisconsin Golden Acre' cabbage and 'Early White' cauliflower, and *B. rapa* 'WR 70 Days' (Figdore et al. 1988) and 'Michihili'(Song et al. 1988a) Chinese cabbage. These clones have been used to infer phylogenetic relationships in *Brassica* species (Song et al. 1988a,b, 1990), assessment of polymorphism in cole crops (Figdore et al. 1988), RFLP mapping (Slocum 1989; Slocum et al. 1990; Song et al. 1990) and tagging of clubroot resistance genes (Landry et al. 1992). A second set of genomic clones derived from *EcoRI* genomic libraries of *B. oleracea* 'Red on Green' flowering kale and rapid cycling *B. napus* (Hosaka et al. 1990). Although 139 low copy clones were selected as potentially useful probes from these libraries, only fourteen have been used for genetic mapping in *B. oleracea* (Hosaka et al. 1990; McGrath et al. 1990; Kianian and Quiros 1992a), *B. rapa* (McGrath and Quiros 1991) and *B. nigra* (This et al. 1990; Chevre et al. 1991). Most of these clones were selected for genome specificity (Hosaka et al. 1990). Recently, another *Pst* I genomic library of *B. napus* 'Westar' was created by Hoenecke and Chyi (1991). Fragments ranging from 500 to 3000 bp were used to generate 450 clones, of which 130 were mapped in *B. napus* and 160 in *B. rapa*.

2.2.1.2. *cDNA clones*. The main source of cDNA clones used to develop probes for RFLP mapping derived from a rapid cycling *B. napus* (CrGC5) embryo library constructed by Harada et al. (1988). Eighteen seedling-specific clones isolated by Harada and his group have been used for mapping in *B. napus* (Landry et al. 1991), *B. oleracea* (McGrath et al. 1990; Kianian and Quiros 1992a) and *B. rapa* (McGrath and Quiros 1991).

This library was made available to several laboratories for the isolation of probes. Landry et al. (1991) isolated 198 anonymous clones ranging from 350 to 2500 bp for the construction of *B. napus* and *B. oleracea* linkage maps. Kianian and Quiros (1992a) isolated 45 clones from 500 to 2500 bp for constructing a *B. oleracea* and a *B. rapa* map (McGrath and Quiros 1991). Also some of these probes have been mapped in *B. nigra* (This et al. 1990; Chevre et al. 1991; Struss et al. 1991b).

2.2.1.3. *Gene specific probes*. Several gene specific probes have been used in *Brassica* RFLP mapping. Sequences from napin (pN2, Crouch et al. 1983) and cruciferin (pC1, Simon et al. 1985) from *B. napus* have been mapped in *B. oleracea* by Song et al. (1988b), McGrath et al. (1990) and Kianian and Quiros (1992b), and in *B. napus* by Landry et al. (1991). pN2 has been mapped in *B. rapa* by McGrath and Quiros (1991) and Schilling and Bernatzky (pers. comm.) and in *B. nigra* (This et al. 1990). Also radish sequences for napin (pB69), cruciferin (pAF7) and late embryogenesis abundant protein (lea, probe p8B6) (Raynal et al. 1991) have been mapped in *B. nigra*. Another 'lea' sequence (pLEA76) from *B. napus* was mapped in *B. rapa* by McGrath and Quiros (1991) and in *B. oleracea* by Kianian and Quiros (1992b). Other *Brassica* gene sequences mapped are: self-incompatibility (BOS5, Nasrallah et al. 1985) in *B. oleracea* (Kianian and Quiros 1992b), malate synthase (pBSMS1, Comai et al.

1989a) and isocitrate lyase (pBSIL9, Comai et al. 1989b) in *B. oleracea* (Kianian and Quiros 1992b) and *B. rapa* (McGrath and Quiros 1991; Schilling and Bernatzky, pers. comm.). The following heterologous sequences have been mapped in *B. oleracea* (McGrath et al. 1990) and *B. rapa* (Schilling and Bernatzky, pers. comm.): alcohol dehydrogenase (kAt3011) from *Arabidopsis thaliana* (Chang and Meyerowitz 1986), chlorophyll binding protein (pL8, Pichersky et al. 1985), ribosomal genes from wheat (pTA71, Gerlach and Bedbrook 1979) or radish (D12-3, Delcasso-Tremousaygue et al. 1988) in *B. oleracea* (McGrath et al. 1990; Delseny et al. 1990; Kianian and Quiros 1992b) and in *B. rapa* (Schilling and Bernatzky, pers. comm.).

2.2.1.4. *Degree of polymorphism.* In agreement with morphological variation, extensive RFLP polymorphism for genomic clones has been detected in *Brassica* diploid cultivated species. This is explained in part by the almost obligate outcrossing of these species imposed by self-incompatibility. Figdore et al. (1988) in a comprehensive study concluded that the high level of polymorphism in *B. oleracea* and *B. rapa* warrant the use of only one restriction enzyme and many clones instead of several enzymes and few clones. Based on 361 anonymous genomic clones from three libraries and three enzymes *EcoRI*, *EcoRV* and *HindIII*, they observed that *B. oleracea* and *B. rapa* were different in 95% of the comparisons. This percentage dropped to 70% for some intraspecific comparisons. A similar situation has been observed for other probes, including anonymous cDNA clones. For example, McGrath and Quiros (1991) found polymorphisms for 69% of the probes in the parents used for generating an F₂ linkage map in *B. rapa* with a single restriction enzyme (*EcoRI*). Kianian and Quiros (1992a) observed also similar levels of polymorphisms in four sets of *B. oleracea* parents used to generate a multipopulation F₂ linkage map with the enzyme *EcoRI*; these ranged from 45% of probes for collard 'Georgia' and cauliflower 'Purple Cauliflower' to 71% in kohlrabi 'Early Purple Vienna' and *B. insularis*. Gene specific sequences seem to display similar levels of polymorphisms. Segregating rDNA intergenic spacer sequences in *B. oleracea* were mapped in three independent chromosomes by Kianian and Quiros (1992b). Also napin, cruciferin, malate synthase, isocitrate lyase and self-incompatibility sequences displayed enough polymorphisms for *EcoRI* digests of parents of four F₂ progenies to allow their mapping. These sequences, except for cruciferin and self-incompatibility, were mapped in a single *B. rapa* family based on *EcoRI* digests. Although in general, the level of polymorphism is high in the diploid cultivated species, it drops to lower levels when broken down to specific chromosomes for some parental combinations. For example, Table 1, extracted from Kianian and Quiros (1992a), illustrates the absence of segregating RFLP markers for chromosome 8 in a collard by cauliflower F₂ family. In such cases, more than one restriction enzyme may be necessary to disclose enough segregating loci for mapping.

RFLP polymorphism for the amphidiploid species is restricted to only two examples in *B. napus*. Landry et al. (1991), in a study based on cDNA clones and

Table 1. Number of RFLP markers for six chromosomes in four F₂ progenies of *B. oleracea* (from Kianian and Quiros 1992a).

Chromosome	col × broc	col × cau	kale × cau	kohl × <i>B.i.</i>	Total
1	9 (60%)	8 (53%)	6 (40%)	8 (53%)	15
2	8 (40%)	2 (10%)	10 (50%)	11 (55%)	20
4	8 (73%)	5 (45%)	4 (37%)	6 (54%)	11
5	6 (42%)	3 (21%)	8 (57%)	4 (29%)	14
6	3 (60%)	3 (60%)	2 (40%)	2 (40%)	5
8	5 (45%)	0 (0%)	4 (37%)	7 (64%)	11

Col = collard, cau = cauliflower, bro = broccoli, kohl = kohlrabi, *B.i.* = *B. insularis*.

digests for the enzymes *Bam*HI, *Eco*RI, *Eco*RV and *Hind*III, found that 59% of the probes detected RFLPs with at least one enzyme. Had they used a single enzyme, the authors estimate that this value may have dropped to 27%. Hoenecke and Chyi (1991) using genomic clones in *Eco*RI digests of a F₂ progeny implied that only approximately 29% of the clones produced RFLPs. Thus, it is apparent that the level of polymorphisms generated by the two different sources of probes is very similar. In any case, the polymorphism of *B. napus* is considerably lower than that observed in *B. oleracea* and *B. rapa*. This is expected because *B. napus* has a lower level of out-crossing due to a weak and often non-existing self-incompatibility system. However, it should be kept in mind that the *B. napus* gene pool sampled by the two studies is fairly narrow since it included only a few rapeseed cultivars.

2.2.2. RAPD markers

We have developed a series of genome specific markers by polymerase chain reaction (PCR) DNA amplification using arbitrary 10-mer oligonucleotides (Quiros et al. 1991), following the technique proposed by Williams et al. (1990). Several of these markers have been assigned to their respective chromosomes in the B genome using *B. napus-nigra* alien addition lines (Quiros et al. 1991). Similarly, we have mapped sequences coding for isocitrate lyase and cruciferin by PCR amplification using specific primers constructed from known sequences of those genes. We have added RAPD markers to our RFLP based maps of *B. oleracea* and *B. nigra*. The great advantages of this technique are its simplicity, rapidity and the minute amounts of DNA necessary for the sample. However, it discloses mostly dominant markers, although several strategies have been proposed to generate co-dominant RFLPs from RAPD markers when necessary (Michelmore et al. 1991).

2.2.2.1. *Degree of polymorphism.* Large amounts of polymorphisms for RAPD markers were reported by Quiros et al. (1991) after surveying 47 arbitrary primers in the three diploid and the three amphidiploid cultivated *Brassica* species. The size of the amplified fragments ranged from 200 to 4500 bp. Three types of markers were detected: 1) Genome-specific bands present in both

diploids and derived amphidiploids sharing the same genome, 2) Species-specific markers unique to accessions of a single species, and 3) Accession-specific markers displayed by some accessions of a particular species, representing intra-specific variability. Half of the primers detected B genome specific markers, supporting the greater divergence of the B genome from the A and C genomes. The three types of markers detected by RAPDs make the technique versatile and applicable to various genetic and evolutionary studies. For example, the genome specific markers serve effectively to identify alien chromosomes in a series of addition lines involving two genomes (Quiros et al. 1991). Species specific markers have been applied to the study of phylogenetic relationships of various *Brassica* species (Truco and Quiros 1991). The intra-specific variability can be used for fingerprinting varieties of various crops and to follow their pedigrees (Hu and Quiros 1991b). Also this type of variability can be exploited to generate linkage maps in segregating progenies. In *B. oleracea* we have estimated that the chances of finding the same marker segregating in two independent progenies is approximately 25% (Hu and Quiros 1991b). Thus, the probability of finding markers already mapped in one population in new families is relatively low. This makes difficult the wide application of pre-mapped RAPD markers on linkage maps to other breeding populations. However, the simplicity of the technique and the availability of excellent programs for linkage map construction (Lander et al. 1989; Stam 1993; Suiter et al. 1983) warrants the development of progeny-specific maps tailored to the genes of interest segregating in those progenies.

3. Maps

Two types of maps are available in *Brassica* – synteny maps based on alien addition lines, and F₂ linkage maps. The integration of these maps is an ongoing effort that we are addressing in our research activities. The conventional nomenclatures of *Brassica* genome and chromosomes (Robbelen 1960) were used for designating the chromosomes and linkage groups.

3.1. Synteny maps

These maps are available for most of the chromosomes of the B and C genomes. Since no addition lines dissecting the A genome are available, this type of map has not been elaborated for this genome.

3.1.1. C genome maps

We have developed two synteny maps for the 9 C-genome chromosomes including isozyme, RFLP and RAPD markers. One map has approximately 194 markers and was constructed using a set of alien addition lines *B. rapa-oleracea* extracted from artificial *B. napus* 'Hakuran' (McGrath and Quiros 1990; McGrath et al. 1990) (Fig. 1). The second map has approximately 104 markers

C1	C2	C3	C4	C5	C6	C7	C8	C9
pINF6	kA3011	○pB185	+pB547	●6pgd-1	pINF9	pAX58-1	pAX58-2	□C09-1300
◆pB370-2	pBN10	○pB850-1	+pBN113	Got-5	p2BN3	pBN11	pB845-2	□C16-1700
pBN127-2	pBN109	○pBN5	+pBN127-1	○pINF10	p2NA9	pBN116	pB850-2	
pC1	pBN111	○pBN7-1	pBN25	pINF8	pB370-1	pBN122-2	pB8MS1-2	
pCOT44-2	pBN120	pD12-3	+pBN63-1	○p2NB3	pBN12	pBN130	pCA3-1	
pLEA76	pBSMS1-1	○pTA71	+pBN79	pBN128-2	pBN121	pBN142	pPBN66	
pN2	pTA71	pYLC24	+pBSIL9-1	○pBN129-2	pBN121	pBN63-2	pYLC2	
◆A06-1500	A03-1000	A04-450	+pLEA76	pBN136	pBN122-1	pBN69	A08-700	
A07-2100	A04-300-	A04-1200	pL8	pBN14	pBN128-1	pBN99	B03-400	
A17-800	A04-450	○A05-400	A01-1900	pBN33	pBN129-1	pBN99-2	C01-1200	
B06-900	A07-1400	A17-700	+A03-450	○pBN46	pBN157	pCA12	C04-1600	
B12-650	A07-2100	B01-1000	A07-2100	○pBN6	○pBN55	pCOT44-1	C12-1100	
C01-850	A17-700	○B05-900	A11-1200	pCA3-2	pBN57	pYLC19	◆C16-520	
C06-510	B06-900	C05-500	B06-900	○pCOT46	pBN98	pYLC5		
C13-850	B07-510	C07-900	B19-550	pL8	pD12-3	A03-1000		
C12-1500	B07-850	○C09-2800	C05-480	PLF	Pgm-1	A07-2100		
C17-480	B19-380	C12-420	C10-450	A01-1400	*pTA71	A17-900		
	B19-1800	C13-1200	+C13-750	●A03-1300	A03-1000	B06-900		
	C07-900	C20-1050	+C14-900	●A09-1000	A04-450	C02-1000		
	*C09-520	D01-1100	C18-750	A11-550	A06-800	C07-1200		
	C15-600	E20-650	C19-100	A11-2400	A07-1400	*C08-400		
	C20-1050	E20-950	C20-480	A17-900	A07-2100	C12-420		
			D01-2500	C01-510	A17-700	*C12-1600		
				C02-1000	*A17-2000	*C14-1500		
				C07-1200	*B05-500	C17-480		
				C08-400	B07-510	C17-3200		
				C08-900	B12-800			
				C14-1500	*C07-900			
				C17-3500	C07-1500			
				C19-700	C12-420			
				C19-1300	C18-600			
				D01-800	C18-650			
					C19-1000			
					C20-1050			
					D06-900			
					E20-1400			

Fig. 1. Nine C genome synteny groups constructed with *B. rapa-oleracea* addition lines derived from 'Hakuran'. Markers in bold labelled by special symbols are those in common with the synteny groups shown in Fig. 2. Asterisk and bar indicate markers whose location has been assigned to either chromosome 8 or 9. Adapted from McGrath et al. (1990).

and was assembled from a set of addition lines *B. rapa-oleracea* extracted from natural *B. napus* (Quiros et al. 1987) (Fig. 2).

The synteny groups of both maps can be aligned by a series of common markers in spite of changes in synteny arrangement for some of the chromosomes which will be discussed below. In addition, a partial synteny map has been constructed from a synthetic *B. napus* obtained after crossing turnip (*B. rapa*) × kohlrabi (*B. oleracea*). This one has 5 synteny groups that align also with the chromosomes of the first two maps (Hu and Quiros, unpub.).

3.1.1.1. *Determination of general stability of C genome alien chromosomes.* We have addressed this problem by following groups of syntenic markers in the progeny of monosomic addition lines (Hu and Quiros 1991a). Data have been obtained from two progenies of approximately 100 plants each, derived from two monosomic addition lines of *B. rapa-oleracea* for chromosomes C4 and C5 ($2n = 21$). After following several markers located on each chromosome, the alien chromosomes were found not to be completely stable. All the expected markers were recovered in approximately 50% of the plants carrying the alien chromosome. The remaining $2n = 21$ plants had one or more markers missing.

These observations were extended to eight chromosomes of the C genome in the 'Hakuran' derived lines with a larger number of markers. This study

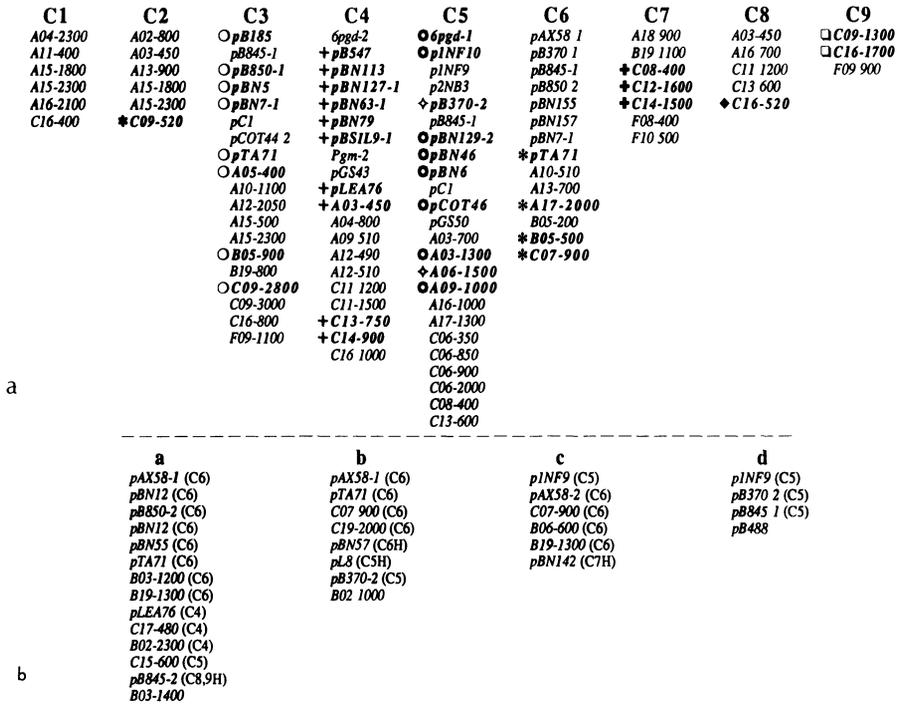


Fig 2 a) Nine C genome synteny groups constructed with *B rapa-oleracea* alien addition lines derived from *B napus* (B430) Markers in bold labelled with special symbols are those in common with the synteny groups shown in Fig 1 Group C5 has markers from 'Hakuran' groups C1 and C5 b) Four recombinant synteny groups (a, b, c, and d) recovered from *B rapa-oleracea* alien addition lines derived from *B napus* (B430) (see Fig 2a) The origin of the markers is shown in parenthesis Those followed by H were observed in the 'Hakuran' derived groups Their absence in the parental *B napus* derived groups may have resulted from chromosome deletions

demonstrated that some of the chromosomes in the C genome are more susceptible to breakage than others. For example, chromosomes C2 and C7 were not recovered intact in any progeny. On the other hand, chromosome C8 did not suffer deletions (Table 2). The deletions were terminal, except for chromosomes C2 and C4 (from natural *B. napus*) which each showed an interstitial deletion. Cytological inspection revealed that the loss of markers was associated with reduced size of the alien chromosome. Therefore, it is possible to map physically the location of the markers on the alien chromosomes by deletion analysis (Hu and Quiros 1991a).

The stability of the alien chromosomes seems adequate for use in gene mapping. However, to assure the integrity of the chromosome in derived progenies it is necessary to monitor the presence of markers along the chromosome.

Table 2. Chromosome deletions observed in progenies of *B. rapa-oleracea* monosomic addition lines.

Chromosome	Number of plants	Number of markers	Number of deletions			% Intact chromosomes
			Terminal	Interstitial	Total	
1	5	10	9	0	9	20
2	4	14	7	1	8	0
3	3	15	1	0	1	67
4	3	14	2	0	2	33
5	2	16	4	0	4	50
6	3	19	2	0	2	33
7	3	12	6	0	6	0
8	4	6	0	0	0	100
Total	27	106	31	1	32	

3.1.2. *B* genome maps

Four partial sets of addition lines from different origins and at different stages of development have been constructed for the B genome. One set is a *B. oleracea-nigra* series extracted from *B. carinata* (Quiros et al. 1986) including only 4 or 5 of the 8 B genome chromosomes. The second set is a *Diplotaxis erucoides-B. nigra* set covering 7 of the 8 chromosomes in the B genome (Quiros et al. 1988; This et al. 1990). The third and fourth sets are in a tetraploid background, *B. napus-nigra*, and were developed independently in France (Jahier et al. 1989, Chevre et al. 1991) and Germany (Struss et al. 1991a,b). Chevre's set is the most extensive, covering at least 7 of the 8 B genome chromosomes. The alien chromosomes are in disomic condition and contain RFLP, RAPD and isozyme markers. Also in this set the chromosomes carrying genes for fatty acid chain elongation and desaturation and possibly a chromosome carrying resistance to blackleg have been identified. The addition lines developed by Struss et al. (1991a) include B genome chromosomes extracted from *B. nigra*, *B. carinata* and *B. juncea*. These represent 6 of the 8 B genome chromosomes with isozyme, RAPD and RFLP markers. Chromosomes carrying erucic acid, sinigrin and possible blackleg resistance have been identified (Struss et al. 1991b).

These lines are being used to assign *B. nigra* linkage groups under construction from two F₂ segregating populations.

3.2. F₂ linkage maps

3.2.1. *B. oleracea* maps

Slocum et al. (1990) reported an extensive RFLP map of 258 markers covering 820 recombination units in 9 linkage groups with average intervals of 3.5 units. This is a proprietary map developed from a broccoli × cabbage cross. Landry

et al. (1992) have constructed a *B. oleracea* map consisting of 201 RFLP markers distributed on 9 major linkage groups covering 1112cM.

Kianian and Quiros (1992a) developed a *B. oleracea* map comprising 108 markers, spread in 11 linkage groups and covering 747 cM. This map was based on three intra-specific and one inter-specific F_2 populations, namely: Collard \times cauliflower, collard \times broccoli, kale \times cauliflower and kohlrabi \times *B. insularis*. The map was constructed using the program Linkage 1 (Suiter et al. 1983) and manual calculations for gene order. The majority of the markers in the map are RFLP loci, with a few morphological and isozyme markers. Recently, we have added a few RAPD markers and redrawn the map based only on the three intra-specific crosses using JOINMAP (Stam et al. 1993) (Fig. 3). It consists of 82 markers distributed on 8 major linkage groups covering 431 cM. The *B. oleracea* linkage groups disclosed were located to their respective chromosomes by using the *B. rapa-oleracea* addition lines. The size of the map was reduced because markers deviating from Mendelian segregation were not included. Also, the elimination of the inter-specific progeny, which showed chromosomal rearrangements, decreased the number of segregating markers in the map.

3.2.2. *B. rapa* maps

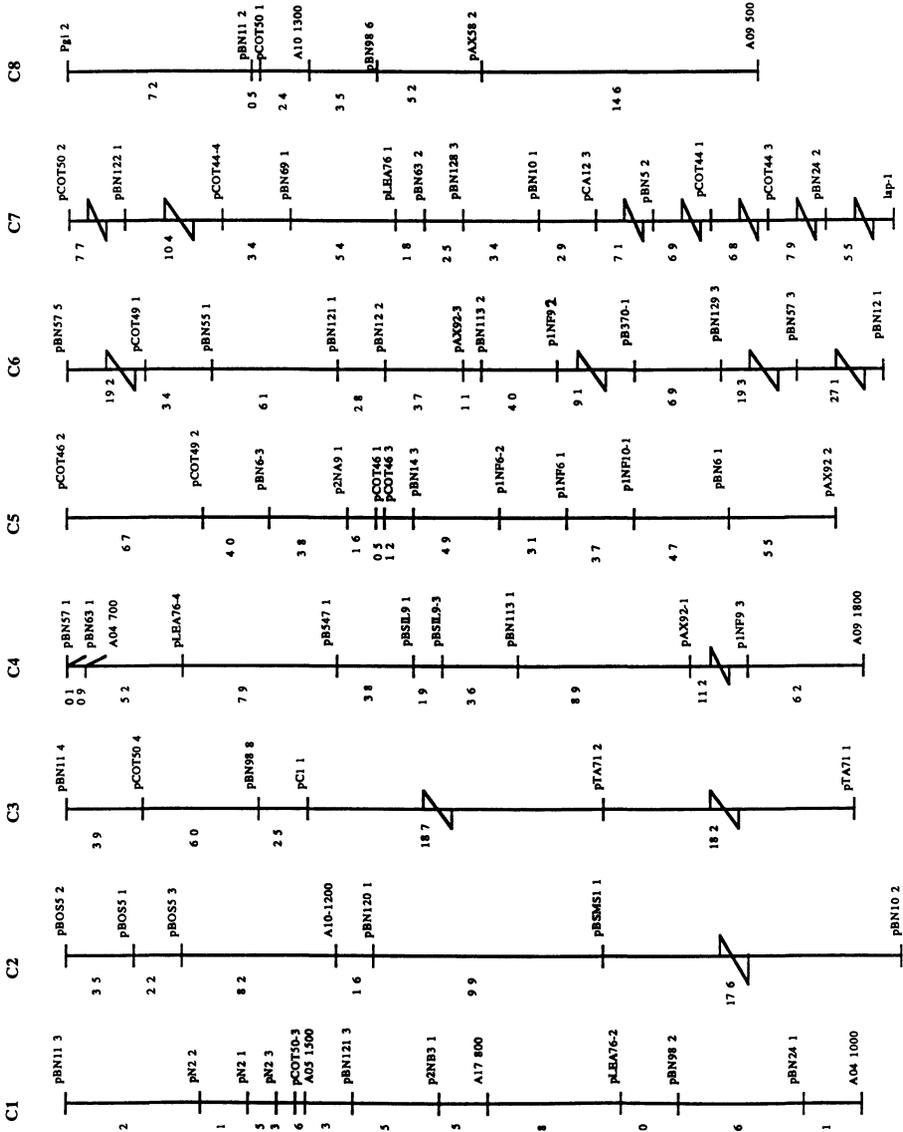
Slocum (1989) and Song et al. (1991) developed a proprietary RFLP map for this species by crossing Chinese cabbage \times 'spring broccoli'. This map includes 273 loci covering 1455 recombination units in 10 linkage groups with average intervals of 5 units. Schilling and Bernatzky (Bernatzky, pers. comm.) constructed an RFLP map of 58 loci covering approximately 700 cM after crossing the oilseed cultivar 'Candle' by a rapid cycling strain.

McGrath and Quiros (1991) developed a third *B. rapa* linkage map from an F_2 of turnip \times Pak Choi containing 49 markers in 8 linkage groups and covering a total of 262 cM (Fig. 4). In addition to molecular markers, this population also segregated for presence of hypocotyl and root enlargement, and wide petioles.

Recently, Chyi et al. (1992) reported a fourth map for this species consisting of 360 RFLP markers distributed on 10 linkage groups covering 1876 recombination units.

3.2.3. *B. nigra* map

This map is based on a single F_2 population of 83 plants, involving two parental individuals from geographically divergent populations (Truco and Quiros 1992). The map constructed with the program Mapmaker (Lander et al. 1988), has 67 markers arranged in 8 major linkage groups which may correspond to the 8 *B. nigra* chromosomes, plus two small groups. Most of the markers are RAPD's with a few isozymes and RFLPs based on *EcoRI* digestions. The map covers 561 cM with average intervals of 8.4 cM. (Fig. 5). Four of the linkage groups have been assigned to their respective chromosomes using alien addition lines (Chevre et al. 1991).



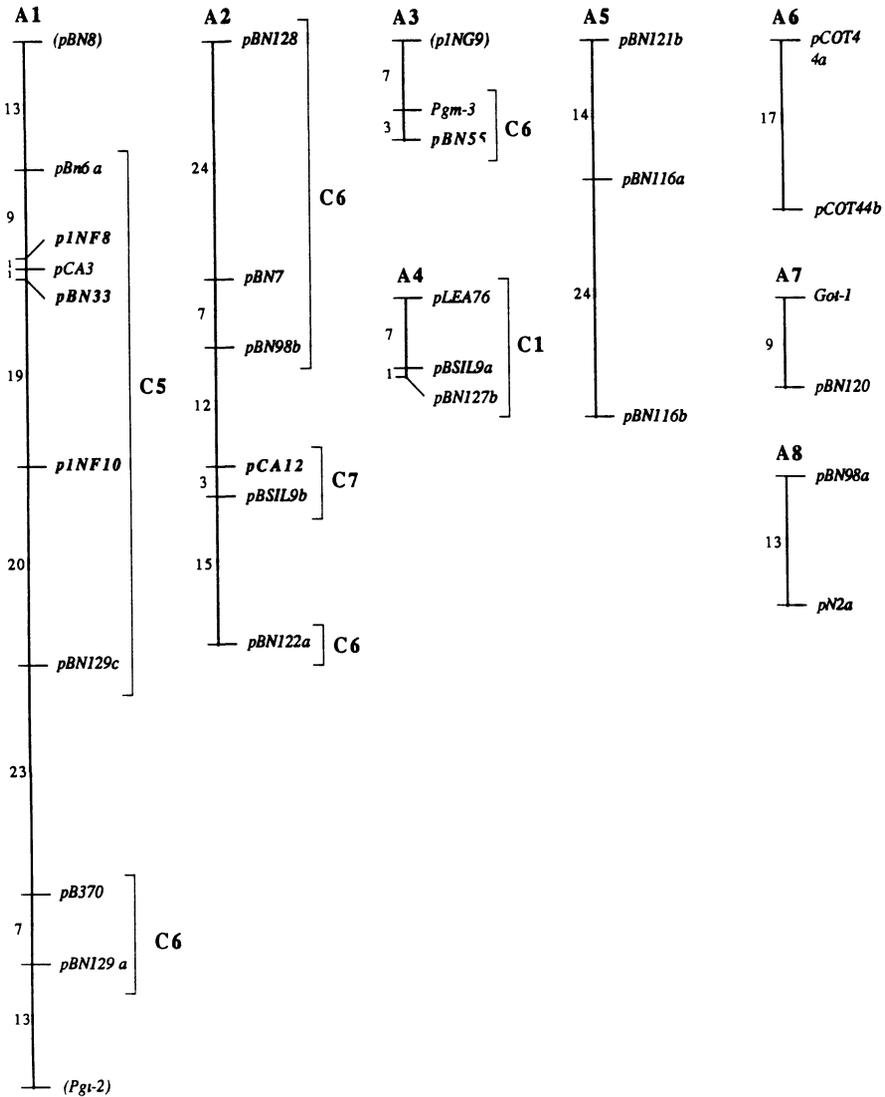


Fig 4 *B rapa* linkage groups (A genome) and its correspondence with *B oleracea* (C genome) synteny groups derived from 'Hakuran'. Adapted from McGrath and Quiros (1991)

←

Fig 3 C genome linkage groups based on three segregating F_2 populations involving the following crosses Collard \times cauliflower, collard \times broccoli and kale \times cauliflower (adapted from Kianian and Quiros 1992a). Linkage between two rDNA loci (pTA71-1 and pTA71-2) in C3 was detected only in collard \times cauliflower progeny. Three loci (pBN24-2, pCOT50-2 and pLEA76-4) of C7 were located in C1 in kale \times cauliflower progeny. These two exceptions may represent rearrangements.

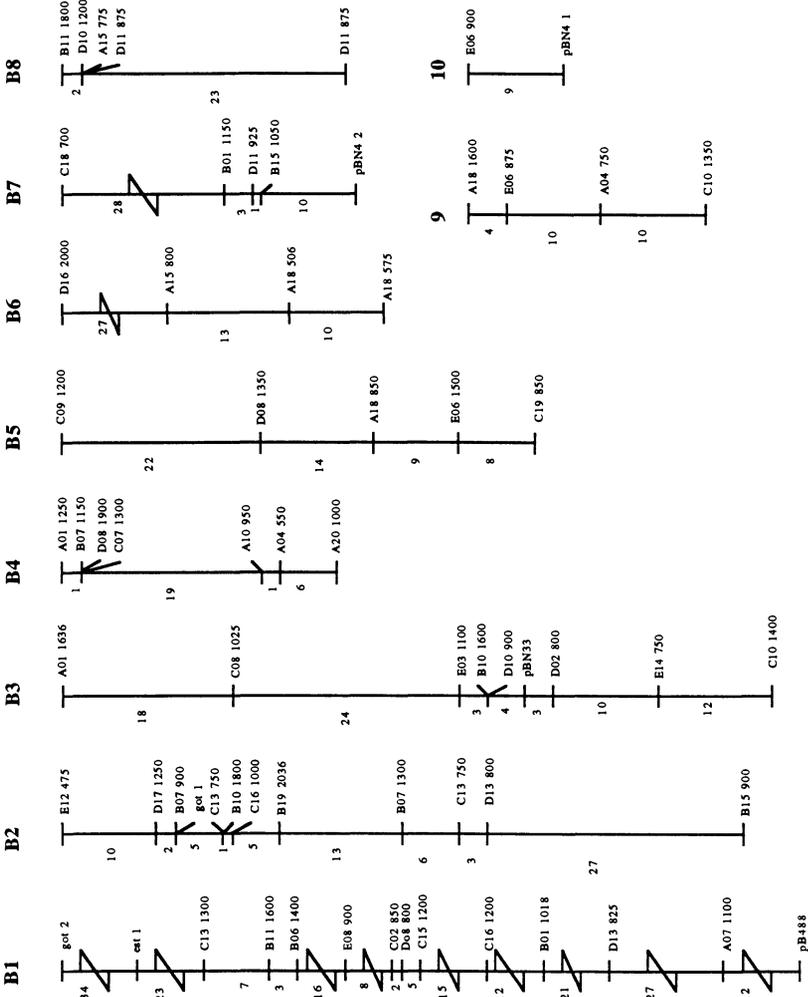


Fig 5 *B. nigra* F₂ linkage map (B genome) based on RAPD, RFLP and isozyme markers

3.2.4. *B. napus* map

Two maps of this amphidiploid species have been reported by independent laboratories (Landry et al. 1991; Hoenecke and Chyi 1991). The first map was developed by crossing two rapeseed cultivars, 'Westar' and 'Topaz', based on single-enzyme digestions by four enzymes, *Bam*HI, *Eco*RI, *Eco*RV and *Hind*III. It includes 120 loci arranged in 19 linkage groups covering 1413 recombination units (Landry et al. 1991). The second is a proprietary map developed by crossing two breeding lines BN0011 × BN0031 and using on *Eco*RI digestions. This map consists of 125 markers arranged in 19 linkage groups covering 1350 map units (Hoenecke and Chyi 1991).

4. Structure of the *Brassica* genomes

4.1. Genomic composition

Most of the information on genomic structure is available for the C and A genomes. Analysis of synteny maps for the C genome (McGrath et al. 1990) (Figs. 1, 2a and b), and the F₂ linkage maps for the C and A genomes (Slocum et al. 1990; Song et al. 1991; McGrath and Quiros 1991; Kianian and Quiros 1992a) (Figs. 3 and 4) reveals extensive sequence duplication. For example, Slocum et al. (1990) reported for *B. oleracea* that 35% of the genomic clones produced more than one locus, and 56% besides disclosing single locus segregations often produced other monomorphic fragments that may represent duplications. Kianian and Quiros (1992a) also in *B. oleracea* found 56% of their cDNA sequences mapping to more than one locus. McGrath et al. (1990) reported over 40% sequence duplications in *B. oleracea* chromosomes represented in a series of addition lines. Song et al. (1991) reported in *B. rapa* that 36% of their genomic clones produced segregating RFLPs at more than one locus and 41% detected sequences segregating as a single locus but also detected additional monomorphic fragments. Thus, these reports demonstrate that approximately 50% of the loci in the genomes C and A are duplicated, supporting the hypothesis that the *Brassica* diploid species are secondary polyploids (Prakash and Hinata 1980). In general these duplications are distributed in more than one chromosome. Some of the linkage and synteny groups have sequences present in two other groups. When linkage arrangement is conserved, most commonly the distances are changed. The rearrangements of linkage groups may be explained by chromosomal translocations. Such aberrations are of common occurrence in *Brassica* and have been reported by independent investigators as a widespread event in various species (Snogerup 1980; Quiros et al. 1988). In addition to duplications and linkage rearrangements, deletions seem to be another important molding force of the *Brassica* genomes. The four independent F₂ linkage studies cited above have detected a large number of loci containing null alleles which may be due to

deletions. Sometimes these loci are assembled in linkage blocks implying large deletions in some chromosomes (McGrath and Quiros 1991; Song et al. 1991). The presence of deletions has been demonstrated cytologically in the C genome using alien addition lines (Hu and Quiros 1991a). Finally, inversions have also been observed in the F₂ linkage maps of *B. oleracea* by Song et al. (1991) and Kianian and Quiros (1992a).

Similar levels of genome duplication have been reported in *B. napus*. Landry et al. (1991) found that 88% of the probes disclosed more than one locus. *B. napus* being an amphidiploid, these duplications are expected to correspond to both intra-genomic and inter-genomic sequences. This indeed seems to be the case, because some of the probes disclosed three or four segregating loci located on different chromosomes (Landry et al. 1991; Hoenecke and Chyi 1991).

Another interesting attribute of the *Brassica* genomes is the relatively large number of loci deviating from Mendelian segregation, and their arrangement in linkage blocks (McGrath and Quiros 1991; Figdore et al. 1992; Landry et al. 1991; Kianian and Quiros 1992a). These deviations may indicate genomic divergence between the parents involved in the crosses (McGrath and Quiros 1991). This conjecture is supported by the fact that the number of deviating loci increase with the divergence of the parents (Kianian and Quiros 1992a).

From the presence of the other chromosomal aberrations in addition to the duplications, the A and C genomes do not have a polysomic structure where specific whole chromosomes are present in two or three copies as implied by formulae based on 6 founder chromosomes (Sikka 1940; Robbelen 1960). Rearranged duplicated segments resulting in novel syntenic combinations seem to be the rule in the C genome. These duplicated segments, however, have undergone sequence changes thus yielding restriction fragments of different sizes when tested by the same probe. In spite of the extensive rearrangements, the ancestry of a few chromosomes can be followed by some of the duplicated loci (Kianian and Quiros 1992a) (Fig. 6). For example, chromosomes 3 and 8 share four duplications suggesting a common origin. It is possible that chromosomal structural modifications, such as reciprocal and overlapping translocations, may have been the main driving force molding the C genome. These are known to yield duplications (Gottlieb 1983), and aneuploids such as tertiary trisomics (Khush and Rick 1967). However, interspecific hybridization between individuals of similar genomes may have also been involved in this process.

Little information is available for other *Brassica* genomes. However, the few isozyme and RFLP loci tested in the B genome (This et al. 1990, Chevre et al. 1991, Struss et al. 1991) show a similar trend. Furthermore, the $n = 7$ genomes of *B. adpressa* (syn. *Hirshfeldia incana*) and *D. eruroides* share most of the duplicated isozyme loci reported for the $n = 8, 9$ and 10 species (Quiros et al. 1988).

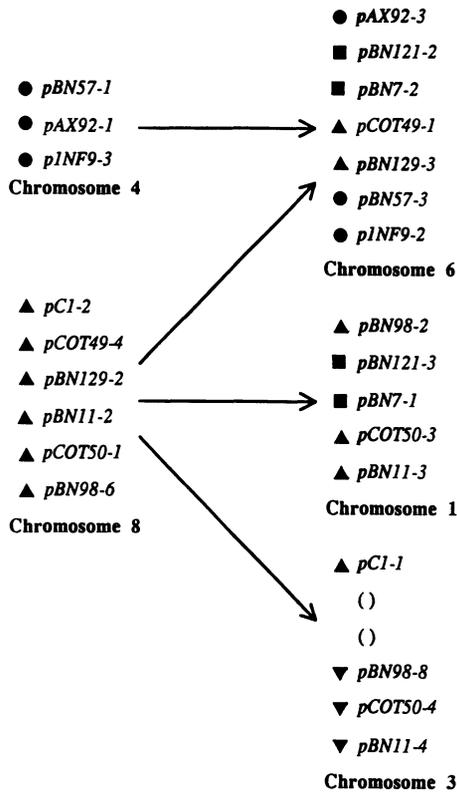


Fig 6 Hypothetical evolutionary pathway of five C genome chromosomes based on duplicated loci disclosed by linkage analysis Sequences of the putative ancestral chromosomes are labelled by symbols () indicate deletions, inverted triangles on chromosome 3 indicates possible inversion involving three markers in the 8 chromosome Adapted from Kianian and Quiros (1992a)

4.2. On the origin of the genomes

Ideally, the use of common probes in segregating progenies of the different diploid species could serve to compare linkage arrangement of RFLP markers. However, this comparison presents several difficulties intrinsic to the duplicated nature of the *Brassica* genome. Most of the probes disclose more than one segregating locus, which makes it difficult to separate orthologous from paralogous loci, thus making genomic comparisons tentative. Based on a limited number of RFLP markers identified in a series of alien addition lines for the C genome which were also segregating in a *B. rapa* progeny, we did a comparative study between the A and C genomes (McGrath and Quiros 1991). We found that the genic complement of both genomes seems similar both at the level of nucleotide sequence and gene copy number, as all tested probes hybridized with both species and no change was detected in copy number or single copy loci. In general, conservation of syntenic chromosome segments was

observed between the two genomes. However, often segments from more than one C genome chromosome were fused together forming part of an A genome linkage group (Fig. 4). Similar observations were reported by Slocum (1989) after comparing F_2 linkage maps of *B. oleracea* and *B. rapa* created by the same probes. She was able to align chromosome C5 with A5 showing strong linkage arrangement conservation. However, some major rearrangements were observed. For example, chromosome C1 shared most of its markers with A1 and a few with chromosome A10. Similarly, chromosome A5 merged markers from both C4 and C5 chromosomes. Again, it can be concluded from these two studies that the difference in structure between the A and C genomes is not simply due to polysomy of an additional ancestral chromosome as specified by the genomic formulae based on the existence of 6 basic chromosomes (Sikka 1940; Robbelen 1960). Most likely, the '10th chromosome' of the A genome is a reshuffled chromosome derived from duplications and rearrangements of the others present in the ancestral genome. Thus, the amount of genetic information contained in the A genome may not be more than that present in the C genome, in spite of having an additional chromosome. This is supported by recent DNA content measurements indicating that the A genome has 468 to 515 Mbp whereas the C genome has 599 to 662 Mbp (Arumuganathan and Earle 1991).

Although the duplicated nature of the *Brassica* genomes supports the hypothesis that the A, B and C genomes derive from smaller ones followed by duplications, it does not provide any evidence of ascending evolution from $n = 6$ to $n = 10$ in a stepwise fashion. Although it is possible that the C genome ($n = 9$) may have contributed to the origin of the A genome ($n = 10$), genomic and organelle molecular data (Warwick and Black 1991; Song et al. 1990) indicate that it is unlikely that the C genome originated directly from the B genome ($n = 8$).

4.3. Plasticity of the *Brassica* genomes

4.3.1. Intragenomic homoeologous recombination

Observations comparing the chromosomes of addition lines derived from natural and synthetic *B. napus* 'Hakuran' dissecting the C genome (McGrath et al. 1990) disclosed possible syntenic differences. Although chromosomes 3, 4, 5 and 6 align well in both sets, chromosomes 1, 2 and 7 display rearrangements of the markers. The C genome chromosomes extracted from 'Hakuran' align fairly well with linkage groups produced by four F_2 *B. oleracea* families. Thus, the C genome chromosomes extracted from natural *B. napus* possibly have suffered rearrangements after amphiploidy. Other changes observed in the two sets of lines are the dispersion of the RFLP marker pB845 and the presence of four rDNA chromosomes in the material derived from natural *B. napus* (Hosaka et al. 1990). A third set of addition lines derived from a kohlrabi (*B. oleracea*) \times turnip (*B. rapa*) amphidiploid is being developed for comparing syntenies with the two previous sets (Quiros et al., unpub.). However, preliminary observations indicate that some of these chromosomes have syntenic

arrangements not observed in the first two sets of addition lines. Evidence from the C-genome-derived addition lines (McGrath et al. 1990) indicates that non-homologous recombination may take place in the *B. oleracea* genome when two or more of these chromosomes are present in single copies as alien additions. Furthermore, in addition to the 9 expected synteny groups resolved from the *B. napus* C genome, four recombinant groups including markers from different chromosomes were recovered (Fig. 2b). The chromosomes most often involved in these recombination events were C4, C5 and C6, indicating that some chromosomes are more recombinogenic than others. Interestingly, two of these chromosomes were found to have enough duplications to follow their 'phylogeny' (Fig. 6) in the F₂ linkage map reported by Kianian and Quiros (1992a). The synteny changes observed in the alien addition lines occurred in 12% of the plants carrying more than one alien chromosome. Although synthesis of interspecific aneuploids for extracting addition lines may promote intragenomic recombination, it does not explain the occurrence of synteny changes in diploid crops of the same species such as those observed by Kianian and Quiros (1992a) in *B. oleracea* crops and the closely related species *B. insularis*.

Variations in synteny have also been observed for the B genome chromosomes in addition lines from independent origins (This et al. 1990; Chevre et al. 1991; Struss et al. 1991b). The information is too limited to ascertain the effect of amphiploidy on these genomic changes. The comparative map for *B. rapa* and *B. napus* (Hoenecke and Chyi 1991) illustrates significant linkage arrangement differences between the A genomes from the diploid and amphidiploid species. However, it is still possible to identify major conserved linkage groups.

The structural changes observed in each of the three cultivated genomes indicate that this phenomenon is widespread and not exclusive of a single genome. Although amphiploidy and interspecific aneuploidy may serve to induce these genomic changes, some rearrangements take place in the diploid environment. Considering the duplicated nature of the genomes and conservation of linkage blocks among chromosomes within the genome, these rearrangements likely are due to intragenomic homoeologous recombination events.

4.3.2. Intergenomic homoeologous recombination

Occasionally a few of the diploid individuals derived from *B. rapa-oleracea* monosomic addition plants were found to carry a few C-genome-specific markers present in the alien chromosomes of the parental plant, indicating that intergenomic recombination had taken place. Earlier during the development of these lines, we detected intergenomic recombination between the A and C genome chromosomes for the isozyme locus *Pgi-2* (Quiros et al. 1987). Recently we have observed intergenomic recombinants for rDNA sequences, where *B. rapa* individuals display *EcoRI* fragments typical of *B. oleracea* (Hu et al., unpub.). Possible intergenomic recombination has also been observed in *B.*

napus-nigra addition lines by Struss et al. (1991b). Another line of evidence for this type of recombination has been obtained by Lydiate et al. (1991) who followed segregation of genome specific RFLP markers in *B. napus* progenies.

5. Applications of the maps in breeding

The *Brassica* linkage maps are starting to be applied to the solution of various breeding problems. For example, Landry et al. (1992) identified in a *B. oleracea* segregating progeny associations between RFLP markers and loci determining clubroot resistance. Landry et al. (1991) are using their *B. napus* map to tag cytoplasmic male-sterility restorer and maintainer genes, glucosinolate content and resistance to blackleg disease. Chyi et al. (1991) used an RFLP map to assess the recovery of the recurrent parent genome in a backcross progeny of *B. rapa*. The DNA-based markers used for the creation of these maps have also been used for variety identification and fingerprinting (Hu and Quiros 1991b).

Although the number of markers for the *Brassica* species is very extensive, they are not consolidated into a single map for each species. However, the rearrangements observed within each genome may make impossible the alignment of some of the chromosomes. Another problem is the use of different restriction enzymes to generate polymorphisms, which will make the alignment of the loci time-consuming. The availability of alien addition lines will serve to assign at least some of the linkage groups to their respective chromosomes, but again their use will be limited only to those few markers which happen to segregate in the mapping progenies. Lastly, another obstacle to the consolidation of these maps is the proprietary nature of many probes and maps.

6. References

- Arumuganathan, K. and Earle, E.D. (1991) Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9: 208–219.
- Arus, P. and Orton, T.J. (1983) Inheritance and linkage relationships of isozyme loci in *Brassica oleracea*. *J. Hered.* 74: 405–412.
- Attia, T. and Robbelen, C. (1986) Cytogenetic relationship within cultivated *Brassica* analyzed in amphihaploids from the three diploid ancestors. *Can. J. Genet. Cytol.* 28: 323–329.
- Attia, T., Busso, C. and Robbelen, G. (1987) Digenomic triploids for an assessment of chromosome relationships in the cultivated diploid *Brassica* species. *Genome* 29: 326–330.
- Banga, S.S. (1988) C-genome chromosome substitution lines in *Brassica juncea* (L.) Coss. *Genetica* 77: 81–84.
- Carr, R.A. and McDonald, B.E. (1991) Rapeseed in a Changing World: Processing and Utilization. GCIRC 8th Int. Rapeseed Congr. 1: 39–56. Saskatoon, Saskatchewan, Canada.
- Chang, C. and Meyerowitz, E.M. (1986) Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. *Proc. Natl. Acad. Sci. U.S.A.* 83: 1408–1412.
- Chang, R.Y., Tai, W. and Fan, Z. (1987) Cytogenetic studies of aneuploids in rapeseed. II. Morphology, fertility, and cytology of self-pollinated progenies of monosomic plants of *Brassica napus*. *Genome* 29: 174–179.
- Chen, B.Y., Heneen, W.K. and Simonsen, V. (1990) Genetics of isozyme loci in *Brassica campestris*

- L and in the progeny of a trigonomic hybrid between *B napus* L and *B campestris* L Genome 33 433-440
- Chevre, A M , This, P , Eber, F , Renard, M , Delseny M and Quiros, C F (1991) Characterization of disomic addition lines of *Brassica napus-Brassica nigra* by isozyme, fatty acids and RFLP markers Theor Appl Genet 81 43-49
- Chyi, Y S , Hoenecke, M and Sernyk, L (1991) Utilization of restriction fragment length polymorphism linkage map in rapeseed breeding In Rapeseed in a Changing World, Proc 8th Int Rapeseed Congr 5 1516-1521 Saskatoon, Saskatchewan, Canada
- Chyi, Y S , Hoenecke, M and Sernyk, L (1992) A genetic linkage map of restriction fragment length polymorphism loci for *Brassica rapa* (syn *campestris*) Genome 35 746-757
- Comai, L , Baden, C S and Harada, J J (1989a) Deduced sequence of a malate synthase polypeptide encoded by a subclass of the gene family J Biol Chem 264 2778-2782
- Comai, L , Dietrich, R A , Maslyar, D J , Baden, C S and Harada, J J (1989b) Coordinate expression of transcriptionally regulated isocitrate lyases and malate synthase genes in *Brassica napus* L Plant Cell 1 293-300
- Coulthart, M and Denford, K E (1982) Isozyme studies in *Brassica* I Electrophoretic techniques for leaf enzymes and comparison of *B napus*, *B campestris* and *B oleracea* using phosphoglucosmutase Can J Plant Sci 62 621-630
- Crouch, M L , Tenbarge, K M , Simon, A E and Ferl, R (1983) cDNA clones from *Brassica napus* seed storage proteins evidence from nucleotide sequence analysis for both subunits of napin are cleaved from a precursor polypeptide J Mol Appl Genet 2 273-283
- Delcasso-Tremousaygue, D , Grellet, F , Panabiers, F , Anaviev, E D and Delseny, M (1988) Structural and transcriptional characterization of the external spacer of a ribosomal RNA nuclear gene from a higher plant Eur J Biochem 172 767-776
- Delseny, M , McGrath, M , This, C , Chevre, A and Quiros, C F (1990) Ribosomal RNA genes in diploid and amphidiploid *Brassica* and related species organization, polymorphism and evolution Genome 33 733-744
- Erickson, L R , Straus, N A and Beversdorf, W D (1983) Restriction patterns reveal origins of chloroplast genomes in *Brassica* amphiploids Theor Appl Genet 65 201-206
- Fan, Z and Tai, W (1985) A cytogenetic study of monosomics in *Brassica* Can J Gen Cytol 27 683-688
- Fan, Z , Tai, W and Stefansson, B R (1985) Male sterility in *Brassica napus* L associated with an extra chromosome Can J Genet Cytol 27 467-471
- Fantes, J A and MacKay, G R (1979) The production of disomic addition lines of *Brassica campestris* Crucifer Newsletter 4 36-37
- Figdore, S S , Kennard, W C , Song, K M , Slocum, M K and Osborn, T C (1988) Assessment of the degree of restriction fragment length polymorphism in *Brassica* Theor Appl Genet 75 833-840
- Gerlach, W L and Bedbrook, J R (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley Nucl Acids Res 7 1869-1885
- Gottlieb, L D (1983) Isozyme number and phylogeny In U Jensen and D E Fairbrothers (eds), Proteins and Nucleic Acids in Plant Systematics Springer-Verlag, Berlin
- Haga, T (1938) Relationship of genome to secondary pairing in *Brassica* (a preliminary note) Jpn J Genet 13 277-284
- Harada, J J , Baden, C S and Comai, L (1988) Spatially regulated genes expressed during seed germination and postgerminative development are activated during embryology Mol Gen Genet 212 466-473
- Hoenecke, M and Chyi, Y S (1991) Comparison of *Brassica napus* and *B rapa* genomes based on restriction fragment length polymorphism mapping In Rapeseed in a Changing World, Proc 8th Int Rapeseed Congr 4 1102-1107 Saskatoon, Saskatchewan, Canada
- Hosaka, K , Kianian, S F , McGrath, J M and Quiros, C F (1990) Development and chromosomal localization of genome specific DNA markers of *Brassica* and the evolution of amphidiploid and $n = 9$ diploid species Genome 33 131-142
- Hu, J and Quiros, C F (1991a) Molecular and cytological evidence of deletions in alien

- chromosomes for two monosomic addition lines of *Brassica campestris-oleracea* Theor Appl Genet 81 221–226
- Hu, J and Quiros, C F (1991b) Identification of broccoli and cauliflower cultivars with RAPD markers Plant Cell Rep 10 505–511
- Jahier, J, Chevre, A M, Tanguy, A M, Eber, F (1989) Extraction of disomic addition lines of *Brassica napus-B nigra* Genome 32 408–413
- Kaneko, Y, Matsuzawa, Y and Sarashima, M (1987) Breeding of the chromosome addition lines of radish with single kale chromosome Jpn J Breed 37 438–452
- Kianian, S F and Quiros, C F (1992a) Generation of a *Brassica oleracea* composite RFLP map linkage arrangements among various populations and evolutionary implications Theor Appl Genet 84 544–554
- Kianian, S F and Quiros, C F (1992b) Genetic analysis of multigene families in *Brassica oleracea* and related species Genome 35 516–527
- Khush, G S and Rick, C M (1967) Tertiary trisomics origin, identification, morphology and use in determining position of centromeres and arm location of markers Can J Genet Cytol 9 610–631
- Landry, B S, Hubert, N, Etoh, T, Harada, J J and Lincoln, S E (1991) A genetic map for *Brassica napus* based on restriction fragment length polymorphisms detected with expressed DNA sequences Genome 34 543–552
- Landry, B S, Hubert, N, Crete, R, Chang, M S, Lincoln, S E and Etoh, T (1992) A genetic map of *Brassica oleracea* based on RFLP markers detected with expressed DNA sequences and mapping of resistance genes to race 2 of *Plasmiodiophora brassicae* (Woronin) Genome 35 409–420
- Lydiate, D, Sharpe, A, Keith, D and Rutier, R (1991) Recombination between homoeologous chromosomes in synthetic *Brassica napus* Int Soc Plant Mol Biol 3rd Int Congr, Tucson, Arizona Poster Abstract – 1720
- McGrath, J M and Quiros, C F (1990) Generation of alien addition lines from synthetic *B napus* morphology, cytology, fertility, and chromosome transmission Genome 33 374–383
- McGrath, J M and Quiros, C F (1991) Inheritance of isozyme and RFLP markers in *Brassica campestris* and comparison with *B oleracea* Theor Appl Genet 82 668–673
- McGrath, J M, Quiros, C F, Harada, J J and Landry, B S (1990) Identification of *Brassica oleracea* monosomic alien chromosome addition lines with molecular markers reveals extensive gene duplication Mol Gen Genet 223 198–204
- Michelmore, R W, Paran, I and Kesseli, R V (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis a rapid method to detect markers in specific genomic regions using segregating populations Proc Natl Acad Sci U S A 88 9828–9832
- Mizushima, U (1950) Karyogenetic studies of species and genus hybrids in the tribe *Brassicaceae* of *Cruciferae* Tohoku J Agr Res 1 1–14
- Morinaga, T (1929) Interspecific hybridization in *Brassica* I The cytology of F₁ hybrids of *B napella* and various other species with 10 chromosomes Cytologia 1 16–27
- Morinaga, T (1934) Interspecific hybridization in *Brassica* VI The cytology of F₁ hybrids of *B juncea* and *B nigra* Cytologia 6 62–67
- Morinaga, T and Fukushima, E (1933) Karyological studies on a spontaneous haploid mutant of *Brassica napella* Cytologia 4 457–460
- Nasrallah, J B, Kao, T H, Goldberg, M L and Nasrallah, M E (1985) A cDNA clone encoding an S-locus specific glycoprotein from *Brassica oleracea* Nature 318 263–267
- Olsson, G and Hagberg, A (1955) Investigations on haploid rape Hereditas 41 227–237
- Palmer, J D, Shield, C R, Cohen, D B and Orton, T J (1983) Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species Theor Appl Genet 65 181–189
- Pichersky, E, Bernatzky, R, Tanksley, S D, Breinderbach, R B, Kausch, A P, Cashmore, A R (1985) Molecular characterization and gene mapping of two clusters of genes encoding chlorophyll a/b-binding proteins in *Lycopersicon esculentum* (tomato) Gene 40 247–258
- Prakash, S and Hinata, K (1980) Taxonomy, cytogenetics and origin of crop *Brassicaceae*, a review Opera Bot 55 1–57

- Quiros, C F , Kiaman, S F , Ochoa, O and Douches, D (1985) Genome evolution in *Brassica* use of molecular markers and cytogenetic stocks *Crucifer Newsletter* 10 21–23
- Quiros, C F , Ochoa, O , Kiaman, S F and Douches, D (1986) Evolutionary trends in *Brassica* gathering evidence from chromosome addition lines *Crucifer Newsletter* 11 22–23
- Quiros, C F , Ochoa, O , Kiaman, S F and Douches, D (1987) Analysis of the *Brassica oleracea* genome by the generation of *B rapa-oleracea* chromosome addition lines characterization by isozymes and rDNA genes *Theor Appl Genet* 74 758–766
- Quiros, C F , Ochoa, O and Douches, D S (1988) Exploring the role of $x = 7$ species in *Brassica* evolution hybridization with *B nigra* and *B oleracea* *J Hered* 79 351–358
- Quiros, C F , Hu, J , This, P , Chevre, A M and Delseny, M (1991) Development and chromosomal localization of genome specific markers by polymerase chain reaction in *Brassica* *Theor Appl Genet* 82 627–632
- Raynal, M , Depigny, D , Grellet, F and Delseny, M (1991) Napin genes characterization of radish cDNA clones and evolution of the gene family in cruciferae *Gene* 99 77–86
- Robbelen, G (1960) Beitrage zur Analyse des *Brassica*-Genoms *Chromosoma* 11 205–228
- Sampson, D R (1966) Genetic analysis of *Brassica oleracea* using nine genes from sprouting broccoli *Can J Genet Cytol* 8 404–413
- Sikka, S M (1940) Cytogenetics of *Brassica* hybrids and species *J Genet* 40 441–509
- Simon, A E , Tenbarge, K M , Scofield, S R , Finkelstein, R R and Crouch, M L (1985) Nucleotide sequence of a cDNA clone of *Brassica napus* 12S storage protein shows homology with legumin from *Pisum sativum* *Plant Mol Biol* 5 191–201
- Slocum, M K (1989) Analyzing the genomic structure of *Brassica* species using RFLP analysis In T Helentjaris and B Burr (eds), *Development and Application of Molecular Markers to Problems in Plant Genetics* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Slocum, M K , Figdore, S S , Kennard, W C , Suzuki, J Y and Osborn, T C (1990) Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea* *Theor Appl Genet* 80 57–64
- Snogerup, S (1980) The wild forms of the *Brassica oleracea* group ($2n = 18$) and their possible relations to the cultivated ones In S Tsonuda, K Hinata and C Gomez-Campo (eds), *Brassica Crops and Wild Allies Biology and Breeding*, pp 121–132 Japan Sci Soc Press
- Song, K M , Osborn, T C and Williams, P H (1988a) *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs) 1 Genome evolution of diploid and amphidiploid species *Theor Appl Genet* 75 784–794
- Song, K M , Osborn, T C and Williams, P H (1988b) *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs) 2 Preliminary analysis of subspecies within *B rapa* (syn *campestris*) and *B oleracea* *Theor Appl Genet* 76 593–600
- Song, K , Osborn, T C and Williams, P H (1990) *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs) 3 Genome relationships in *Brassica* and related genera and the origin of *B oleracea* and *B rapa* (syn *campestris*) *Theor Appl Genet* 79 497–506
- Song, K M , Suzuki, J Y , Slocum, M K , Williams, P H and Osborn, T C (1991) A linkage map of *Brassica rapa* (syn *campestris*) based on restriction fragment length polymorphism loci *Theor Appl Genet* 82 296–304
- Stam, P (1993) Construction of integrated genetic linkage maps by means of a new computer package *Join Map The Plant J* 3 739–744
- Struss, D , Bellin, U and Robbelen, G (1991a) Development of B-genome chromosome addition lines of *B napus* using different interspecific *Brassica* hybrids *Plant Breed* 106 209–214
- Struss, D , Quiros, C F and Robbelen, G (1991b) Construction of different B-genome addition lines of *Brassica napus* L In *Rapeseed in a Changing World Breeding Biotechnology GCIRC 8th Int Rapeseed Congr Saskatoon* 2 358–362
- Suiter, K A , Wendel, J F and Case, J S (1983) Linkage-1 a pascal computer program for the detection and analysis of genetic linkage *J Hered* 74 203–204
- This, P , Ochoa, O and Quiros, C F (1990) Dissection of the *Brassica nigra* genome by chromosome addition lines *Plant Breed* 105 211–220
- Truco, M J and Quiros, C F (1991) Evolutionary study on *Brassica nigra* and related species In

- Rapeseed in a Changing World: Breeding: Biotechnology. GCIRC 8th Int. Rapeseed Congr. Saskatoon 2: 318–323.
- Truco, M.J. and Quiros, C.F. (1992) Construction of a linkage map in *B. nigra*. Proc. Seventh Crucifer Genetics Workshop. Corvallis, Oregon, March 1992.
- U, N. (1935) Genome-analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Jpn. J. Genet. 7: 389–452.
- Vaughan, J.G. (1977) A multidisciplinary study of the taxonomy and origin of *Brassica* crops. BioScience 27: 35–40.
- Warwick, S.I. and Black, L.D. (1991) Molecular systematics of *Brassica* and allied genera (subtribe *Brassicinae*, *Brassicaceae*) – chloroplast genome and cytodeme congruence. Theor. Appl. Genet. 82: 81–92.
- Williams, J.G., Kubelik, A.E., Levak, K.J., Rafalski, J.A. and Tingey, S.C. (1991) DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res. 18: 6531–6535.

12. Genetic mapping in lettuce

RICHARD W. MICHELMORE¹, RICHARD V. KESSELI² and EDWARD J. RYDER³

¹ Department of Vegetable Crops, University of California, Davis, CA 95616, U.S.A.; ² Department of Biology, University of Massachusetts, Harbor Campus, Boston, MA 02125-3393, U.S.A.;

³ USDA,ARS, U.S. Agricultural Research Station, 1636 East Alisal, Salinas, CA 93905, U.S.A.

Contents

1. Introduction	223	3.4. SCARs	230
2. Classical markers	224	4. The genetic map	230
3. Molecular markers	228	5. Applications and future prospects	234
3.1. Isozymes	228	6. Acknowledgements	236
3.2. RFLPs	229	7. References	236
3.3. RAPDs	230		

Abbreviations: LOD – log of the odds ratio (log of the likelihood of the data arising due to linkage over the likelihood of the data arising by chance); NIL – near-isogenic line; PCR – polymerase chain reaction; QTL – quantitative trait locus; RFLP – restriction fragment length polymorphism; RAPD – random amplified polymorphic DNA; SCAR – sequence characterized amplified region

1. Introduction

Cultivated lettuce (*Lactuca sativa* L.) is a diploid ($2n = 18$) species in the Cichoreae tribe of the Compositae (Asteraceae) family. There are three well-established wild species in the subsection *Lactuca*, *L. serriola*, *L. saligna*, and *L. virosa*; all are $2n = 18$ and self-fertilizing. These species can be crossed to *L. sativa* with increasing difficulty in the order listed. Several other sexually compatible species have been described (Ferakova 1977) but their validity as distinct species remains unclear. *L. serriola* is closely related to *L. sativa* and may be conspecific (Lindqvist 1960a; Kesseli et al. 1991). These wild species, especially *L. serriola*, have been sources of several disease resistance genes (see below; Crute 1988); however, they remain a rich source of variation that has yet to be accessed systematically.

Lettuce is amenable to genetic analysis. Most cultivars are highly inbred and exhibit extensive genetic homozygosity. Each plant produces many composite capitulae ('flowers'), each comprising up to 18 true florets. Crosses can be

readily made (Ryder 1986); however, complete outcrossing cannot be guaranteed. Therefore, most genetic analysis of lettuce has relied on F₂ rather than backcross or testcross populations. Each plant produces large numbers of seeds (1,500 to 5,000+) allowing rare recombinants and mutants to be detected. Two to five generations are possible per year depending on the genotype and the environmental conditions. It is also amenable to transgenic analysis. Foreign DNA can be readily introduced into lettuce by cocultivation with *Agrobacterium tumefaciens* (Michelmore et al. 1987), although some genotypes are easier to transform than others.

Lettuce is an important vegetable crop species. The farm gate value in the U.S.A. in 1991 was nearly \$900 million (Anon 1991). It is grown on over 200,000 acres in the U.S.A.. Much of the genetic knowledge on lettuce has been derived from breeding programs particularly in the U.S.A., England, and The Netherlands. Consequently, genetic studies have emphasized disease resistance genes and morphological variation. Recently, there has been increasing emphasis on developing molecular markers as tools to aid in analysis of disease resistance and in selection strategies. This chapter reviews the progress in lettuce genetics over the last ten years.

2. Classical markers

The first genes reported for lettuce were for seed color, anthocyanin production, and lobed leaf (Durst 1930). Since then, a variety of morphological, developmental, and disease and herbicide resistance variants have been identified and characterized genetically. The 59 loci that had been described by 1983 were collated by Robinson et al. (1983) who reviewed and revised the gene nomenclature for lettuce. Since that time, 15 additional classical markers have been identified and named; these include seven disease resistance genes, three leaf color genes, two morphology genes, one flower color gene, and two developmental genes. Table 1 provides an updated list of all genes now known in lettuce. Brief descriptions of the genes that have been identified since 1983 follow.

Table 1. A list of the known qualitative genes for lettuce and related species, 1992.^a

Symbol	Name	Phenotype	Reference
<i>A</i>	<i>Anthocyanin</i>	Purple stem	Ernst-Schwarzenbach (1936)
<i>ag</i>	<i>apple green</i>	Apple green leaf	Ryder (1989)
<i>all,2,3</i>	<i>albino-1,2,3</i>	Albino seedling, lethal	Ryder (1971, 1975)
<i>ax</i>	<i>alboxantha</i>	Chlorophyll deficient	Whitaker (1944)
<i>bi</i>	<i>bidens</i>	Resistance to <i>Bidens</i> mottle virus	Zitter and Guzman (1977)
<i>C</i>	<i>Chirnogen</i>	Complementary gene for expression of anthocyanin	Durst (1930); Thompson (1938)
<i>ca</i>	<i>capitate</i>	Head formation	Lindqvist (1960)

Table 1 Continued

Symbol	Name	Phenotype	Reference
<i>cdl</i> ,2	<i>chlorophyll deficient-1,2</i>	Chlorophyll deficient	Ryder (1975, 1983)
<i>cl</i>	<i>calico</i>	Calico leaf pattern	Whitaker (1968)
<i>cor</i>	<i>corky root resistance</i>	Resistance to corky root rot	Brown and Michelmore (1988)
<i>Cr</i>	<i>Crinkled</i>	Crinkled leaf	Ryder (1965)
<i>ct</i>	<i>cut</i>	Cut, indented margins	Ryder (1965)
<i>Dm</i> #	<i>Downy mildew-</i>	Resistance to downy mildew # = 1,2,3,4,5/8,6,7,10, 11,13,14,15,16	Zink and Duffus (1970, Johnson et al (1977, 1978, Farrara et al (1987)
<i>Efl</i> ,2	<i>Early flowering-1,2</i>	Early flowering	Ryder (1983, 1988)
<i>en</i>	<i>endive</i>	Endive-like leaves	Ryder (1975)
<i>er</i>	<i>erect</i>	Erect involucre, seed non-shattering	Whitaker and McCollum (1954)
<i>fr</i>	<i>fringe</i>	Fringed, twisted leaves	Ryder (1965)
<i>G</i>	<i>green</i>	Complementary gene for expression of anthocyanin	Durst (1930), Thompson (1938)
<i>gl</i>	<i>glossy</i>	Glossy leaf	Lindqvist (1960)
<i>go</i>	<i>golden</i>	Golden flower, chlorophyll deficient	Ryder (1971)
<i>gy</i>	<i>golden yellow</i>	Golden yellow leaves	Bremer (1931)
<i>H</i>	<i>Hearting</i>	Head formation	Lindqvist (1960)
<i>i</i>	<i>intensifier</i>	Intensifies anthocyanin	Lindqvist (1960)
<i>k</i>	<i>kopfbildung</i>	Head formation	Bremer and Grana (1935), Pearson (1956)
<i>lg</i>	<i>light green</i>	Light green leaves	Thompson (1938)
<i>lh</i>	<i>leaf hairs</i>	Leaf hairs, sterility	Ryder (1971)
<i>ms</i> 1,2,3	<i>male sterile-1,2,3</i>	Complementary male sterility genes	Lindqvist (1960)
<i>ms</i> 4,5	<i>male sterile-4 5</i>	Male sterility, epistatic	Ryder (1963)
<i>ms</i> 6	<i>male sterile 6</i>	Male sterility, recessive	Ryder (1967)
<i>Ms</i> 7	<i>Male sterile</i>	Male sterility, dominant	Ryder (1971)
<i>mo</i>	<i>mosaic</i>	Mosaic resistance	Bannerot et al (1969), Ryder (1970)
<i>Nl</i> 1,2	<i>Non-lethal-1,2</i>	Non-lethal mosaic reaction	Zink et al (1973)
<i>P</i>	<i>Pointed</i>	Pointed leaf apex	Lindqvist (1960)
<i>pa</i>	<i>pale</i>	Pale yellow flower	Ryder (1971)
<i>pl</i>	<i>plump</i>	Plump involucre	Ryder (1971)
<i>plr</i>	<i>resistance to Plasmopara lactucae-radicis</i>	Resistance to root downy mildew	Kesseli et al (1993b)
<i>Pm</i>	<i>Powdery mildew</i>	Resistance to powdery mildew	Whitaker and Pryor (1941)
<i>R</i>	<i>Red</i>	Red leaves	Thompson (1938)
<i>Rbs</i>	<i>Red-brown spotted</i>	Red-brown leaf spots	Lindqvist (1958)
<i>Rs</i>	<i>Red-tinged</i>	Red tinged leaf margins	Thompson (1938)
<i>s</i>	<i>smooth</i>	Hairless rib and stem	Durst (1930)
<i>sa</i>	<i>salmon</i>	Salmon flower	Ryder (1989)
<i>Sc</i>	<i>Scalloped</i>	Scallop leaf margin	Ryder (1965)
<i>sg</i>	<i>shiny green</i>	Shiny green leaf	Ryder (1989)
<i>sh</i>	<i>shallow</i>	Shallow corolla cleft	Ryder (1963)
<i>si</i>	<i>sickly</i>	Sickly plant	Ryder (1992)

Table 1 Continued

Symbol	Name	Phenotype	Reference
<i>sn</i>	<i>stunt</i>	Stunted, dark green plants	Ryder (1965)
<i>st</i>	<i>striate</i>	Striate vein pattern	Whitaker and Bohn (1953)
<i>t</i>	<i>tagneutral</i>	Day neutral, slow seed formation, long days	Bremer (1931), Bremer and Grana (1935)
<i>tn</i>	<i>truncate</i>	Truncate leaf	Ryder (1992)
<i>tr</i>	<i>triforine</i>	Toxic reaction to triforine	Maxon-Smith (1979)
<i>Tu</i>	<i>Turnip mosaic</i>	Resistance to Turnip mosaic	Zink and Duffus (1970)
<i>u</i>	<i>unlobed</i>	Unlobed leaves	Durst (1929, 1930)
<i>uo</i>	<i>oak leaf</i>	Oakleaf lobing	Whitaker (1950), Durst (1929)
<i>v</i>	<i>vanishing</i>	Vanishing anthocyanin	Lindqvist (1960)
<i>vi</i>	<i>virescent</i>	Virescent leaf	Ryder (1971)
<i>w</i>	<i>white</i>	White seed	Lindqvist (1960)
<i>y</i>	<i>yellow</i>	Yellow seed	Durst (1929)

^a This is an updated version of a table presented in Robinson et al (1983)

Apple green, ag: The leaves of PI391601, a stem lettuce originating from China, are a pale dull green color that is similar to some varieties of apple. The color value on the Royal Society Colour Chart is 138B, compared to the dull dark green color of cv. Salinas that has a color value of 143C. These colors are easily distinguished at the cotyledon stage and in the late rosette and heading stages in field plantings. At seedling and early rosette stages in the field and most stages in the greenhouse, the colors are less easily distinguished. Apple green is recessive to dull dark green. This difference is conferred by a recessive allele at a single locus which was named *apple green* and designated *ag* (Ryder 1989).

Chlorophyll deficiency, cd-2. This phenotype appeared in a plant of cv. Calmar. When grown from seed in the greenhouse, the cotyledons are light green with a yellowish cast. The first true leaves are yellow-green with a green band along the midvein. Later developing leaves have patches of green and yellow-green. With age, the leaves become markedly yellow with greenish bands along the veins. Chlorophyll deficiency is recessive and due to a single gene. It was named *chlorophyll deficient-2* and designated *cd2* (Ryder 1983).

Corky root resistance, cor: Resistance to corky root rot was first identified in *L. serriola* (Dickson 1963). This was subsequently transferred to several crisphead cultivars, notably Marquette, Montello, and Green Lake (Sequiera 1978). Subsequent genetic studies demonstrated that resistance to the corky root bacterium was conferred by a recessive allele at a single locus, designated *cor* (Brown and Michelmore 1988). Allelism tests indicated that resistance was determined at the same locus in most of the resistant accessions tested.

Downy mildew resistance, Dm1, Dm11, Dm13, Dm14, Dm15, Dm16. Resistance to downy mildew, caused by *Bremia lactucae*, is determined by a gene-for-gene interaction (Crute and Johnson 1976; Illott et al. 1989). Prior to 1983, seven *Dm* genes had been characterized genetically and several other

genetically undefined resistance factors identified (Johnson et al. 1977, 1988). An additional six *Dm* genes have now been identified that confer usually complete resistance to isolates of *B. lactucae* expressing the corresponding avirulence gene (Hulbert and Michelmore 1985; Farrara et al. 1987). *Dm1* was identified in cv. Lednicky. *Dm11* was identified in an experimental line derived from *L. serriola* and in cv. Capitan. *Dm13* was identified in cvs. Pennlake and Vanguard (other *Dm* genes are also present in Vanguard). *Dm14* was identified in cv. Gelber Winterkonig. *Dm15* was identified in *L. serriola*, PIVT1309. *Dm16* was identified in *L. serriola*, LSE/18, and *L. sativa*, cv. Saffier. Linkage analysis demonstrated that all of the 13 *Dm* genes currently known are clustered into four linkage groups along with genes for resistance to other diseases. The largest group contains seven downy mildew resistance genes and a gene for resistance to root aphid. Resistance can be assayed at the cotyledon stage or in mature plants. Additional sources of resistance exist (e.g. Farrara and Michelmore 1987; Bonnier et al. 1992) but have yet to be characterized genetically. Field resistance has also been identified but has not been characterized genetically (Crute and Norwood 1981; Norwood et al. 1985).

Early flowering, Ef1 and Ef2: Two genes were identified in crisphead breeding material that reduced the time to first flower to as few as 45 days. Lettuce cultivars normally require 110 to 140 days under summer greenhouse conditions to progress from seed germination through the rosette and head stages (some cultivars), to the time of first flower. The early flowering phenotype is conditioned by two partially dominant loci, designated *Early flowering-1 (Ef1)* and *Early flowering-2 (Ef2)* (Ryder 1983, 1988). The double homozygous dominant genotype, *Ef1Ef1 Ef2Ef2*, produces only four to five leaves before stem elongation occurs and flowers in 45 days. The genotype, *Ef1Ef1 ef2ef2*, produces a rosette of 10 to 11 leaves before stem elongation and flowers in 65 days. The genotypes, *ef1ef1 Ef2Ef2* and *ef1ef1 ef2ef2* go through normal rosette/head formation before flowering in 100 and 140 days respectively. The earliest flowering genotypes have low seed production and a compacted panicle. Seed production and panicle formation are normal in later flowering genotypes. The *Ef* genes have been used for two purposes. In breeding, they can be used to reduce the generation time by half during introgression of a trait by backcrossing (Ryder 1985, and unpub.). They have also been useful in genetic and developmental studies of stem elongation (Waycott 1989).

Salmon, sa: A new orange-toned variation in flower color was identified in the butterhead cv. Golden Bibb (Ryder 1989). Lettuce flowers are usually yellow. Variations in flower color have previously been described as pale and golden. The orange-toned phenotype is determined by a recessive allele at a single locus; this was named *salmon* and designated *sa*. When *pale* and *salmon* are expressed together, the resultant flower color is pale salmon (Ryder 1992).

Shiny green, sg: This leaf color phenotype is characteristic of the crisphead cultivar Amaral 78, which has glossy or shiny green leaves; this color extends to the elongated stem and to the involucre bracts. This is in contrast to the dull green of cv. Salinas and similar cultivars. The difference in color is particularly

noticeable at the beginning stage of bolting, when the leaves of the latter develop a bluish cast while cv. Amaral 78 remains bright green. It resembles 143B on the color chart. *Shiny green* is due to a recessive allele at a single locus, designated *sg* (Ryder 1989).

Sickly, si: The sickly phenotype was first noticed in a F₂ progeny from a cross between two crisphead types. Seeds of sickly plants germinate normally but the plants grow at a slower rate and produce smaller leaves than normal. Cotyledons are chlorotic with patches of necrotic tissue. The trait is partially lethal with a mortality of about 25%. Surviving plants flower normally and are fertile. The trait is recessive and inherited as a single gene. It was named *sickly* and designated *si* (Ryder 1992).

Truncated leaf, tn: This morphological variant was identified in a crisphead breeding line. Plants with this trait grow normally until approximately the 15 leaf stage, when lateral leaf expansion continues normally but the area around the apex grows more slowly so that the leaf becomes foreshortened. The result is a stunted plant that is darker green than normal. The trait is determined by a recessive allele at a single locus. It was named *truncated leaf* and designated *tn* (Ryder 1992).

3. Molecular markers

3.1. Isozymes

Isozyme variation has been comprehensively surveyed in cultivated lettuce and related wild species. However, isozyme markers have not been used extensively in lettuce beyond these initial studies. Allozyme variation as detected by starch gel electrophoresis was scored among 31 collections of *L. sativa* and wild *Lactuca* spp. and used to assess genetic diversities and phylogenies (Kesseli and Michelmore 1986). Forty-two isozyme systems revealed 70 putative loci; polymorphisms were detected at 22 of these loci. Segregation was demonstrated in F₂ populations for 11 isozyme loci (Kesseli and Michelmore 1986). These were: two of the four loci for alcohol dehydrogenase, *Adh3* and *Adh4*; one of four loci for diaphorase, *Dia3*; five of seven esterase loci, *Est1*, *Est3*, *Est5*, *Est6*, and *Est8*; two of the four loci for glutamate oxaloacetate transaminase, *Got3*, and *Got4*; and one of three loci for malic enzyme, *Me2*. Subsequently, one locus for triose phosphate isomerase (see below, Paran et al. 1992) and one locus for phosphoglucose isomerase (Ochoa and Kesseli, unpub.) have also been characterized genetically. Seven of these loci (*Adh3*, *Dia3*, *Est8*, *Est6*, *Got2*, *Pgi2*, *Tpi2*) have currently been located on the detailed genetic map. Variation was also detected for one of four loci for acid phosphatase, one of two loci for aldolase, one of three loci for fructose 1,6 diphosphatase, one of two loci for glyceraldehyde 3-phosphate dehydrogenase (NAD), one locus for isocitrate dehydrogenase, one of two loci for leucine amino peptidase, one of three loci for malate dehydrogenase, one of two for phosphoglucose isomerase, and one locus for shikimate dehydrogenase. Segregation analysis has yet to be reported for these loci.

Isozyme variation in lettuce has also been studied using polyacrylamide gel electrophoresis (Cole et al. 1991). Thirty accessions of wild *Lactuca* species were assayed for variation using ten enzyme systems. A total of 140 allozyme bands were observed, of which 75 were polymorphic (scored as presence or absence). No genetic studies were reported to identify isozyme loci and alternate alleles. Cluster analyses were conducted to identify banding patterns characteristic of the wild *Lactuca* species and to look for correlations with resistance to root aphid.

A detailed analysis of triose phosphate isomerase has been made at the isozyme and RFLP level (Paran et al. 1992). Several random cDNA clones were identified that detected polymorphism between near-isogenic lines differing for *Dm1* and *Dm3*. One of these clones (pCL1795) detected a multigene family. Subsequent RFLP analysis revealed that other members of this multigene family were linked to the *Dm4* cluster of resistance genes. Sequence analysis of the cDNA clone uncovered similarity to sequences encoding the glycolytic enzyme, triose phosphate isomerase (TPI) in other species. The single locus encoding the cytosolic isozyme was mapped using an inter-specific cross UC83US1 (*L. saligna*) × cv. Vanguard (*L. sativa*); this cosegregated with a RFLP detected by pCL1795. PCR and Southern analysis indicated that the multigene family was the result of amplification of genomic regions because intron sequences had been preserved. This amplification seems to have been recent as related wild *Lactuca* species had many fewer copies. The significance of the linkage between TPI-related sequences and resistance genes is unclear; however, it supports the hypothesis that resistance genes are members of multigene families that have diverged to have different specificities (Michelmore et al. 1987; Pryor 1987).

3.2. RFLPs

Both random cDNA clones and genomic clones have been used as probes for RFLPs in lettuce. A cDNA library was made from *L. serriola*, PI221936 (Landry et al. 1987a). Several different libraries have been used as sources for genomic clones. Some libraries were generated using methylation insensitive enzymes, *MboI*, *BamHI*, *EcoRI* or *MspI*. Others were constructed using methylation sensitive enzymes, *PstI* or *HpaII*. Clones from all genomic libraries were hybridized to labelled total genomic DNA to detect those clones that contained repeated sequences. Only those clones that showed no detectable hybridization were subsequently used as probes for polymorphism. The methylation insensitive enzymes yielded high frequencies of clones containing high copy DNA (average 93%). The frequency was reduced to approximately 50% using methylation sensitive enzymes (Kesseli et al. 1994). Both cDNA and genomic clones detected approximately equal levels of polymorphism (Kesseli et al. 1994). Comparison of the frequency of polymorphism in digests of nine different endonucleases showed no increased polymorphism correlated by the presence of a CpG dimer in the recognition sequence of the restriction enzyme. Digests with enzymes recognizing four base pairs exhibited less frequent

polymorphism than digests with enzymes recognizing six base pairs (Landry et al. 1987a). Most (approximately 70%) of the RFLPs in lettuce seem to be due to insertion/deletion events rather than point mutations at enzyme recognition sites as determined by the correlation of fragment length with polymorphism in multiple enzyme digests (Kesseli et al. 1994). Over 1900 random cDNA clones have now been screened for their ability to detect polymorphism between the parents of our basic mapping population (see below). This has provided 143 RFLP loci on the genetic map. Screening additional cDNA clones was becoming increasingly inefficient due to the presence of duplicate clones. Approximately 400 genomic clones have been screened on the same parents and have provided nine RFLP loci.

3.3. *RAPDs*

RAPD markers were detected in lettuce using procedures similar to Williams et al. (1990) (Kesseli et al. 1994). We have screened over 200 random 10-mer primers (Operon Technologies, Alameda, CA) for their ability to detect segregating loci in our basic mapping population. RAPD loci are no more polymorphic than RFLP loci. An average of 8 bands (potential loci) were observed per primer. Due to the sensitivity of the RAPDs to reaction conditions, RAPD analysis resulted in slightly lower levels of scorable polymorphisms per locus compared with RFLP analysis (Kesseli et al. 1994). However, as large numbers of RAPD loci can be assayed quickly, we are using RAPD markers extensively. Over 250 RAPD loci have now been mapped.

3.4. *SCARs*

Although RAPDs are useful markers for experimental purposes, the sensitivity of RAPD banding patterns to reaction conditions makes them less useful markers for routine analysis of large numbers of plants in breeding and other applications. We have converted RAPD markers in regions of interest to Sequence Characterized Amplified Region (SCAR) markers (Paran and Michelmore 1993). RAPD fragments were cloned and sequenced to provide longer primers (24-mers) for PCR. At high annealing temperatures, these primers amplify only a single locus and are relatively insensitive to reaction conditions. Also, in several cases alleles were amplified from both parents in the basic population so that a dominant RAPD locus was converted into a codominant SCAR locus.

4. The genetic map

Prior to 1983, only four small linkage groups involving ten loci had been reported (Robinson et al. 1983). Extensive intercrosses between lines carrying one or a few markers to develop a genetic map based on classical markers had

not been analyzed. The advent of molecular markers provided the opportunity to develop a detailed genetic map from a single segregating population.

A detailed genetic map of lettuce is being developed from the analysis of a single F_2 population from cv. Calmar \times cv. Kordaat. The basic mapping population comprises 66 F_2 plants and F_3 families derived from them. To obtain more precise gene orders in regions of interest, the mapping population has been increased to 309 individuals. Recombinant inbred lines are being developed from the original 66 F_2 plants. DNA was analyzed from F_2 plants and their F_3 derivatives. Classical markers exhibiting dominance, such as resistance genes, were scored in F_3 families to determine the progenitor F_2 genotype precisely. The linkage relationships were determined using the program, MAPMAKER (Lander et al. 1987). For early maps, a LOD score of 3.0 was used as the threshold for considering linkage significant; this was increased to at least 3.5 in later maps as the number of markers increased. Gene orders were considered correct if the LOD score compared to the next most likely gene order exceeded 2.0. Recombination frequencies were converted to centiMorgans using the mapping function of Kosambi (1944).

An intra-specific cross was utilized so that the resultant genetic map would have direct utility to intraspecific analyses and breeding programs. Cvs. Calmar and Kordaat were chosen as they represent the two most important cultivated plant types of lettuce, crisphead and butterhead, respectively, and they maximize the amount of intra-specific variation. This cross also segregated for six resistance genes (five *Dm* genes and *Tu*). The intra-specific nature of the cross slowed the development of the map but has yielded markers of greater utility than if an interspecific cross had been used. Also, the lines derived from the mapping population can be propagated by seed. Few loci (approximately 9%) exhibited significant segregation distortion making segregation analysis simpler. Regions of repressed recombination due to structural hybridity were not expected.

The first extensive map comprised 53 loci in nine linkage groups of more than three markers (Landry et al. 1987b). These were predominantly RFLP markers, with seven disease resistance, four isozyme, and three morphological loci. This map included 404 cM within the markers. The next map published contained 160 markers arranged in 12 linkage groups of six or more markers and four smaller groups (Kesseli et al. 1990). Eleven loci were unlinked. Apparently 1404 cM were contained within the markers.

The genetic map of lettuce in February 1992 comprised 319 loci, including 152 RFLP, 130 RAPD, 7 isozyme, 19 disease resistance, and 11 morphological markers (Fig. 1; Kesseli et al. 1993a, 1994). Thirteen major and four minor linkage groups have been identified. Nine loci remain unlinked. Nearly 1200 cM are contained within the markers. The decrease in genetic size relative to the previous map resulted from careful checking, rescoring and, if necessary, correction of the data by searching for apparent crossovers on both sides of any marker using the GENOTYPE command of MAPMAKER. A detailed analysis and discussion of the latest map is described in Kesseli et al. (1994).

In order to increase the number of markers in specific regions, we have utilized either near-isogenic lines (NILs), when available, or bulks of individuals from segregating populations. NILs differing for resistance to downy mildew were screened for differences using either pools of random cDNA clones or RAPD primers (Paran et al. 1991). Over 500 cDNA clones were screened to identify four RFLP loci that were subsequently shown to be linked to the *Dm* genes. Two hundred and twelve primers identified 10 RAPD loci linked to *Dm* genes. NILs are not available for most regions of the lettuce genome; therefore, we developed bulked segregant analysis as a method for identifying markers in any region of the genome (Michelmore et al. 1991). This approach was demonstrated by bulking individuals from the basic mapping population of contrasting genotypes for *Dm5/8*. Seventeen homozygous individuals were used to make each bulk. One hundred arbitrary 10-mer primers were screened and three RAPD differences were detected between the bulks. These were then shown to be linked to *Dm5/8* by conventional segregation analysis of the progenitor population. Bulked segregant analysis allows the use of only two samples to identify markers linked to the gene or region on which the bulks are based. It is therefore a rapid method for adding markers to specific regions and for placing classical markers on the map. We have used bulked segregant analysis to help coalesce the map (Kesseli et al. 1994) and quickly to map genes that do not segregate in the basic mapping population (see below).

The map has yet to coalesce into nine linkage groups representing the nine chromosomes. From the data available in 1990, we estimated that doubling the number of markers would saturate the map. This assumed that detectable polymorphic loci were evenly distributed over the genetic map. The failure to complete the map with over 300 markers indicates that this assumption is invalid. This could be due to regions of the genome with greatly reduced polymorphism between Calmar and Kordaat, or due to highly recombinagenic regions. We are attempting to complete the map by several methods. We are continuing to map random RAPD loci to increase the number of loci. We will analyze the segregation of loci at the end of linkage groups on the expanded progeny; this will reduce the standard error of the recombination frequency and therefore allow us to detect linkage over greater genetic distances. We are targeting markers to the end of linkage groups using bulked segregant analysis (Michelmore et al. 1991); this has already brought some linkage groups together (Kesseli et al. 1994). We are also analyzing additional crosses, particularly inter-specific crosses to increase the probability of detecting polymorphic loci. Finally, we will use telomere specific probes to help define the end of linkage groups.

→

Fig. 1. The genetic map of *Lactuca sativa*, Feb. 1992. CL- and GL-prefixes indicate RFLP loci detected by random cDNA and genomic clones respectively OP- and R-prefixes indicate RAPD loci. SC- indicates SCAR loci. Designations of other loci are as in Table 1 except loci that have yet to be described formally: *Ra*, resistance to root aphid; *Cotc* and *Cots*, loci controlling cotyledon color and shape respectively; *Blt*, locus controlling bolting; *Ant*, resistance to anthracnose (Kesseli et al., unpub.; Ochoa et al., unpub.). Different loci in multigene families detected by the same

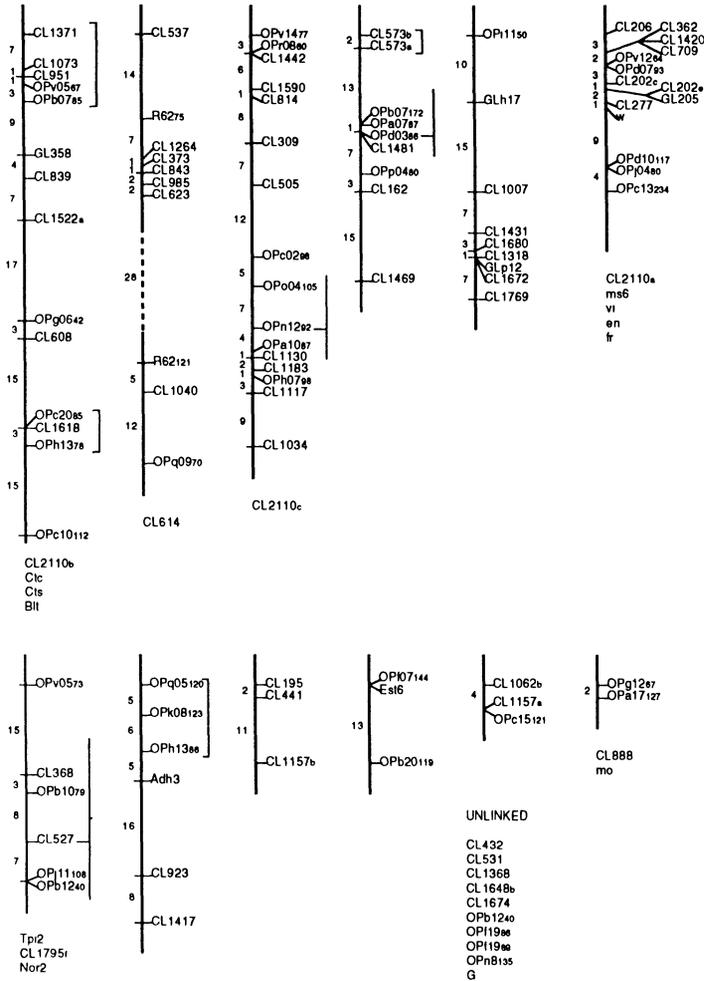


Fig. 1. Continued.

5. Applications and future prospects

We are continuing to develop a detailed map of lettuce. Additional markers will be predominantly PCR-based. RFLP markers remain anchor points on the map due to their codominant nature. Several additional intra- and inter-specific crosses are now being analyzed. These are being used to locate genes that do not segregate in the basic mapping population and to determine to what extent the

map developed from a single F_2 population is representative of the genetic map for *L. sativa* and related species. Particular emphasis is being placed on disease resistance genes. The aim is to map all known genes for disease resistance in lettuce and to develop reliable PCR-based markers to aid in their selection.

We are using bulked segregant analysis to map additional genes quickly. Only six of the genes listed in Table 1 have been placed precisely on the genetic map. The approximate positions are known for another 17 (Fig. 1). Bulks are made for the contrasting genotypes from a population that is segregating for the gene to be mapped. These two bulks are then screened for RAPD differences. When a RAPD is detected, the degree of linkage is determined by conventional segregation analysis of the population from which the bulks were made. If the linked RAPD has not been mapped earlier, it is mapped relative to existing markers using a previously analyzed population, preferably the basic mapping population. When the genetic position of the linked marker is known, other markers in the region are mapped relative to the target gene in the original population to determine the target gene's precise genetic position.

Most of the known disease resistance genes in lettuce will be located on the genetic map. As resistance genes tend to be clustered, mapping resistance genes relative to each other will indicate when there is a danger of introducing susceptibility to one disease while breeding for resistance to another. Multiple SCARs will be developed that flank each resistance gene. This will allow marker-aided selection for resistances that are difficult to score and reduction of the amount of genetic hitch-hiking (linkage drag) occurring during backcross programs. It will also allow the rapid generation of lines carrying multiple disease resistance.

Isozyme and RFLP markers have already been used for phylogenetic studies. Isozyme studies revealed low but detectable intra-line variation (Kesseli et al. 1986). The phylogenies determined from allozyme and RFLP data generally corroborated those based on morphological characters (Kesseli and Michelmore 1986; Kesseli et al. 1991). Accessions that were incorrectly named in the germplasm collection, were obvious from their isozyme and RFLP genotypes. Accessions of each wild species generally clustered as distinct entities. The RFLP data indicated that, while *L. sativa* and *L. serriola* were closely related, *L. sativa* was not a direct extraction from the eight sampled populations of *L. serriola*. Butterhead, crisphead, romaine-like cultivars formed distinct clusters. Numerous genetic differences characterized each plant type indicating polyphyletic origin for *L. sativa*. Future phylogenetic and syntenic studies will likely involve other members of the Compositae using markers developed for *L. sativa*. At present, lettuce is the only member of this very successful family that has been mapped in detail.

Quantitative traits in lettuce have yet to be dissected into their component QTLs. Several horticultural traits, such as plant morphology and speed of bolting, should be amenable to analysis. Another interesting target is field resistance to downy mildew. This seems to be a polygenic trait controlled by relatively few genes (Norwood et al. 1983). Evaluation is complicated by

epistatic effects of plant morphology and maturity on the level of resistance. This prevents the analysis of crosses between markedly different plant types (e.g. butterhead × crisphead); however, crosses between genetically diverse parents are necessary so that sufficient numbers of markers will be polymorphic throughout the genome.

A major effort currently underway is to isolate *Dm* genes by map-based cloning. This involves chromosome walking from tightly linked markers (Michelmore et al. 1992). We are currently saturating regions containing *Dm* genes with markers and determining the relationship between genetic and physical distance in the three main clusters of *Dm* genes. When a marker that is physically close to a targeted *Dm* gene has been identified, we will obtain overlapping genomic clones from a variety of libraries. One of the challenges is to integrate hybridization and PCR-based techniques to allow analysis of these complex regions.

6. Acknowledgements

We acknowledge the numerous people who have contributed over the past ten years to the identification of genes and the development of the map in lettuce, particularly (in alphabetical order): Philip Brown, Barry Farrara, Scot Hulbert, Benoit Landry, Hei Leung, Kiyotaka Nakahara, Oswaldo Ochoa, Ilan Paran, Wei-Ching Wang, and William Waycott. This work has been funded in part by grants, 85-CRCR-1-1561, 88-CRCR-37262-3522 and 91-37300-6364 to RWM from USDA CRGO/NRI and funding from the California Iceberg Lettuce Research Board to EJR and RWM.

7. References

- Anon. (1992) Marketing lettuce from Salinas, Watsonville, other Central California districts and Colorado, 1991. Federal State Market News Service, 28 pp.
- Bannerot, H., Boulidard, L., Marrou, J. and Duteil, M. (1969) Étude de l'hérédité de la tolérance au virus de la mosaïque de la laitue chez la variété Gallega de Invierno. *Ann. Phytopathol.* 1: 219–226.
- Bonnier, F.J.K., Reinink, K. and Groenwold, R. (1992) A search for new sources of major gene resistance in *Lactuca* to *Bremia lactucae*. *Euphytica* 61: 203–211.
- Bremer, A.H. (1931) Einfluss der Tageslänge auf die Wachstumsphasen des Salats. *Genetische Untersuchungen I. Gartenbauwissenschaft* 4: 469–483.
- Bremer, A.H. and Grana, J. (1935) *Genetische Untersuchungen mit Salat. II. Gartenbauwissenschaft* 9: 231–245.
- Brown, P.R. and Michelmore, R.W. (1988) The genetics of corky root resistance in lettuce. *Phytopathology* 78: 1145–1150.
- Cole, R.A., Sutherland, R.A. and Riggall, W.E. (1991) The use of polyacrylamide gel electrophoresis to identify variation in isozymes as markers for *Lactuca* species and resistance to the lettuce root aphid *Pemphigus bursarius*. *Euphytica* 56: 237–242.
- Crute, I.R. (1988) The impact of breeding on pest and disease control in lettuce. *Aspects Appl. Biol.* 17: 305–312.

- Crute, I R and Johnson, A G (1976) The genetic relationship between races of *Bremia lactucae* and cultivars of *Lactuca sativa* Ann Appl Biol 83 125-137
- Crute, I R and Norwood, J M (1981) The identification and characteristics of field resistance to lettuce downy mildew (*Bremia lactucae* Regel) Euphytica 30 707-717
- Durst, C E (1929) Inheritance in lettuce Science 69 553-554
- Durst, C E (1930) Inheritance in lettuce Illinois Agric Exp Sta Bull 356 237-341
- Ernst-Schwarzenbach, M (1936) Fertilitat, Photoperiodismus, und Genetik von *Lactuca sativa* L Zuchter 8 11-21
- Farrara, B and Michelmore, R W (1987) Identification of novel sources of downy mildew resistance in *Lactuca* spp HortScience 22 647-649
- Farrara, B, Ilott, T W and Michelmore, R W (1987) Genetic analysis of factors for resistance to downy mildew (*Bremia lactucae*) in species of lettuce (*Lactuca sativa* and *L. serriola*) Plant Pathol 36 499-514
- Ferakova, V (1977) The genus *Lactuca* L in Europe Univerzita Komenskeho, p 68
- Hulbert, S H and Michelmore, R W (1985) Linkage analysis of genes for resistance to downy mildew (*Bremia lactucae*) in lettuce (*Lactuca sativa*) Theor Appl Genet 70 520-528
- Ilott, T W, Hulbert, S H and Michelmore, R W (1989) Genetic analysis of the gene-for-gene interaction between lettuce (*Lactuca sativa*) and *Bremia lactucae* Phytopathology 79 888-897
- Johnson, A G, Crute, I R and Gordon, P L (1977) The genetics of race specific resistance in lettuce to downy mildew (*Bremia lactucae*) Ann Appl Biol 86 87-103
- Johnson, A G, Laxton, S A, Crute, I R, Gordon, P L and Norwood, J M (1978) Further work on the genetics of race specific resistance in lettuce (*Lactuca sativa*) to downy mildew (*Bremia lactucae*) Ann Appl Biol 89 257-264
- Kesseli, R V and Michelmore, R W (1986) Genetic variation and phylogenies from isozyme markers in species of *Lactuca* Heredity 77 324-331
- Kesseli, R W, Paran, I and Michelmore, R W (1989) *Lactuca sativa* In S J O'Brien (ed), Genetic Maps, 5th ed, pp 6100-6102 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- Kesseli, R V, Ochoa, O and Michelmore, R W (1991) Variation at RFLP loci in *Lactuca* spp and origin of cultivated lettuce Genome 34 430-436
- Kesseli, R V, Paran, I, Ochoa, O, Wang, W -C and Michelmore, R W (1993a) Linkage map of lettuce (*Lactuca sativa*) In S J O'Brien (ed), Genetic Maps, 6th ed , pp 6229-6233 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Kesseli, R V, Paran, I and Michelmore, R W (1994) Analysis of a detailed genetic linkage map of *Lactuca sativa* (lettuce) constructed from RFLP and RAPD markers Genetics 136 1435-1446
- Kesseli, R V, Witsenboer, H, Vandemark, G J, Stangellini, M E and Michelmore, R W (1993b) Recessive resistance to *Plasmopara lactucae-radix* maps by bulked segregant analysis to a cluster of dominant resistance genes in lettuce Mol Plant Microbe Interact 6 722-728
- Kosambi (1944) The estimation of map distance from recombination values Ann Eugen 12 172-175
- Lander, E S, Green, P, Abrahamson, J, Barlow, A, Daly, M J, Lincoln, S E and Newburg, L (1987) MAPMAKER in interactive computer package for constructing primary genetic linkage maps of experimental and natural populations Genomics 1 174-181
- Landry, B S, Kesseli, R V, Farrara, B and Michelmore, R W (1987b) A genetic map of lettuce (*Lactuca sativa* L) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers Genetics 116 331-337
- Landry, B S, Kesseli, R V, Leung, H and Michelmore, R W (1987a) Comparison of restriction endonucleases and sources of probes for their efficiency in detecting restriction fragment length polymorphisms in lettuce (*Lactuca sativa*) Theor Appl Genet 74 646-653
- Lindqvist, K (1958) Inheritance of lobed leaf form in *Lactuca* Hereditas 44 347-377
- Lindqvist, K (1960a) On the origin of cultivated lettuce Hereditas 46 319-350
- Lindqvist, K (1960b) Inheritance studies in lettuce Hereditas 46 387-470
- Maxon-Smith, J W (1979) Triforine sensitivity in lettuce Euphytica 28 351-359
- Michelmore, R W, Marsh, E, Seely, S and Landry, B S (1987a) Transformation of lettuce mediated by *Agrobacterium tumefaciens* Plant Cell Rep 6 439-442

- Michelmore, R W , Hulbert, S H , Landry, B S and Leung, H (1987b) Towards a molecular understanding of lettuce downy mildew In P R Day and G J Jellis (eds), Genetics and Plant Pathogenesis, pp 221–232 Blackwell Scientific, Oxford
- Michelmore, R W , Paran, I and Kesseli, R V (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis a rapid method to detect markers in specific genomic regions using segregating populations Proc Natl Acad Sci U S A 88 9828–9832
- Michelmore, R W , Kesseli, R V , Francis, D M , Paran, I , Fortin, M G and Yang, C -H (1992) Strategies for cloning plant disease resistance genes In S Gurr (ed), Molecular Plant Pathology – A Practical Approach, Vol 2, pp 233–287 IRL Press, Oxford
- Norwood, J M , Crute, I R , Johnson, A G and Gordon, P L (1983) A demonstration of the inheritance of field resistance to lettuce downy mildew (*Bremia lactucae* Regel) in progeny derived from cv Grand Rapids Euphytica 32 161–170
- Norwood, J M , Johnson, A G , O'Brien, M and Crute, I R (1985) The inheritance of field resistance to lettuce downy mildew (*Bremia lactucae*) in the cross 'Avoncrisp' × 'Iceberg' Z Pflanzenzucht 94 259–262
- Paran, I and Michelmore, R W (1993) Identification of reliable PCR-based markers linked to disease resistance genes in lettuce Theor Appl Genet 85 085–993
- Paran, I , Kesseli, R V and Michelmore, R W (1991) Identification of RFLP and RAPD markers linked to downy mildew resistance genes in lettuce using near-isogenic lines Genome 34 1021–1027
- Paran, I , Kesseli, R V and Michelmore, R W (1992) Recent amplification of triose phosphate isomerase related sequences in lettuce Genome 35 627–635
- Pearson, O H (1956) The nature of the rogue in 456 lettuce Proc Am Soc Hort Sci 68 270–278
- Pryor, T (1987) Origin and structure of disease resistance genes in plants Trends Genet 3 157–161
- Robinson, R W , McCreight, J D and Ryder, E J (1983) The genes of lettuce and closely related species Plant Breed Rev 1 267–293
- Ryder, E J (1963a) A gene for depth of corolla cleft in the lettuce flower Veg Imp Newsletter 5 5–6
- Ryder, E J (1963b) An epistatically controlled pollen sterile in lettuce (*Lactuca sativa* L.) Proc Am Soc Hort Sci 83 585–589
- Ryder, E J (1965) The inheritance of five leaf characters in lettuce (*Lactuca sativa* L.) Proc Am Soc Hort Sci 86 457–461
- Ryder, E J (1967) A recessive male sterility gene in lettuce (*Lactuca sativa* L.) Proc Am Soc Hort Sci 91 366–368
- Ryder, E J (1970) Inheritance resistance to common lettuce mosaic J Am Soc Hort Sci 95 378–379
- Ryder, E J (1971) Genetic studies in lettuce (*Lactuca sativa* L.) J Am Soc Hort Sci 96 826–828
- Ryder, E J (1975) Linkage and inheritance in lettuce (*Lactuca sativa* L.) J Am Soc Hort Sci 100 346–349
- Ryder, E J (1983) Inheritance, linkage and gene interaction studies in lettuce J Am Soc Hort Sci 108 985–991
- Ryder, E J (1985) The use of early flowering genes to reduce generation time in backcrossing, with specific application to lettuce breeding J Am Soc Hort Sci 110 570–573
- Ryder, E J (1986) Lettuce breeding In M J Bassett (ed), Breeding Vegetable Crops, pp 433–474 AVI Publishing, Westport
- Ryder, E J (1988) Early flowering in lettuce as influenced by a second flowering time gene and seasonal variation J Am Soc Hort Sci 113 456–460
- Ryder, E J (1989) Studies of three new genes, linkage, and epistasis in lettuce J Am Soc Hort Sci 114 129–133
- Ryder, E J (1992) Lettuce genetics inheritance, linkage, and epistasis J Am Soc Hort Sci 117 (in press)
- Sequra, L (1978) Two root rot resistant varieties of lettuce Wis Agric Exp Sta Rep 2 p
- Thompson, R C (1938) Genetic relations of some color factors in lettuce U S Dept Agr Tech Bul 620

- Waycott, W (1989) Genetic and physiological studies on stem elongation in lettuce (*Lactuca sativa* L.) Ph D thesis, University of California, Santa Cruz
- Whitaker, T W (1944) The inheritance of chlorophyll deficiencies in cultivated lettuce J Hered 35 317-320
- Whitaker, T W (1950) The genetics of leaf form in cultivated lettuce I The inheritance of lobing Proc Am Soc Hort Sci 56 389-394
- Whitaker, T W (1968) A chlorophyll-deficient mutant in lettuce Veg Improv Newsletter 10 5
- Whitaker, T W and Bohn, G W (1953) The striate-vein character in lettuce J Hered 44 177-180
- Whitaker, T W and McCollum, G D (1954) Shattering in lettuce - its inheritance and biological significance Bul Torrey Bot Club 81 104-110
- Whitaker, T W and Pryor, D E (1941) The inheritance of resistance to powdery mildew (*Erysiphe cichoracearum*) in lettuce Phytopathology 31 534-540
- Williams, J G K., Kubelik, A R., Livak, K J., Rafalski, A and Tingey, S V (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers Nucl Acids Res 18 6531-6535
- Zink, F W and Duffus, J E (1970) Linkage of turnip mosaic virus susceptibility and downy mildew, *Bremia lactucae*, resistance in lettuce J Am Soc Hort Sci 95 420-422
- Zink, F W and Duffus, J E (1973) Inheritance and linkage of turnip mosaic virus and downy mildew (*Bremia lactucae*) reaction in *Lactuca serriola* J Am Soc Hort Sci 98 49-51
- Zink, F W Duffus, J E and Kimble, K A (1973) Relationship of a non-lethal reaction to a virulent isolate of lettuce mosaic virus and turnip mosaic susceptibility in lettuce J Am Soc Hort Sci 98 41-45
- Zitter, T A and Guzman, V L (1977) Evaluation of cos lettuce crosses, endive cultivars, and *Cichorium* introductions for resistance to bidens mottle virus Plant Dis Rep 61 767-770

13. RFLP maps of maize

EDWARD H. COE and JACK M. GARDINER

Plant Genetics Research Unit, ARS-USDA, and Department of Agronomy, University of Missouri, Columbia, MO 65211, U.S.A.

Four molecular maps of the maize genome, with Restriction Fragment Length Polymorphisms (RFLPs), have been presented in documented form (Beavis and Grant 1991; Burr and Burr 1991; Gardiner et al. 1993; J. Shoemaker et al., pers. comm. in *Maize Genetics Cooperation News Letter* 66: 65–69, 1992). Combining these maps, two from the public sector and two from the private sector, into a joint, composite map awaits the programming of suitable mapping engines and electronic availability of the scoring data for each (planning and discussions are in progress toward this goal). Previous maps (Burr et al. 1988 et seq.; Helentjaris et al. 1986a,b et seq.) have been subsumed in, or superseded by, the current maps. Table 1 lists the progression and status of maize RFLP maps, with the parentage of the test progeny, progeny type and size, number of probes and probed loci, number of trait loci, and estimated genome size. The proximate joint order of markers, mapped in the public mapping projects, is presented in Fig. 1, using the Core Map developed at the University of Missouri as the framework. The Core Map has been developed to be a statistically qualified standard; regions where the available data for the Core Map leave the order uncertain, because the statistical requirements (a LOD score difference of 3) are not met, are marked with a dotted line. On the whole, the order given agrees with data for the other maps, but locus order in the dotted regions is particularly subject to revision.

Table 1 Molecular-marker maps of maize

Map	Era	Pedigree	Progeny type ^a	Progeny size	Probes no	Probed loci no	Trait ^b loci no	Genome size (cM)
NPI-1	1986	H427/761	F2	50	-	112	5	-
NPI-Composite	1986ff	(Incorporated in BNL)	-	-	-	-	-	-
BNL-1	1988ff	T232/CM37	R1	48	-	533	24	-
BNL-2	1988ff	CO159/Tx303	R1	41	-	698	23	-
BNL-Composite	1992	-	R1	89	-	802	40	2000
UMC-1	1988	Tx303/CO159	F2	46	-	302	0	-
UMC-Core-Composite	1992	Tx303/CO159	IF2	56	190	199	15	1900
Agri genetics-a	1988	A619/Mangelsdorf	F2	99	291	293	6	1800
Agri genetics-b	1988	B68Ht/B73Htrhm	F2	93	87	89	0	1200
Agri genetics-c	1992	De811/B73Htrhm	R1	200	115	120	0	1300
Agri genetics-Composite	1992	-	-	392	314	354	10	-
Pioneer-1	1988	B73/Mo17	F2	112	-	106	0	2100
Pioneer-2	1988	B73/G35	F2	112	-	143	6	2200
Pioneer-3	1988	K05/W65	F2	144	-	78	0	1600
Pioneer-4	1988	J40/V94	F2	144	-	68	0	1500
Pioneer-Composite	1991	-	F2	512	-	196	6	2300

^a R1, Recombinant Inbred population, F2, selfed population, IF2, immortalized F2 population^b Morphological or isozyme polymorphisms

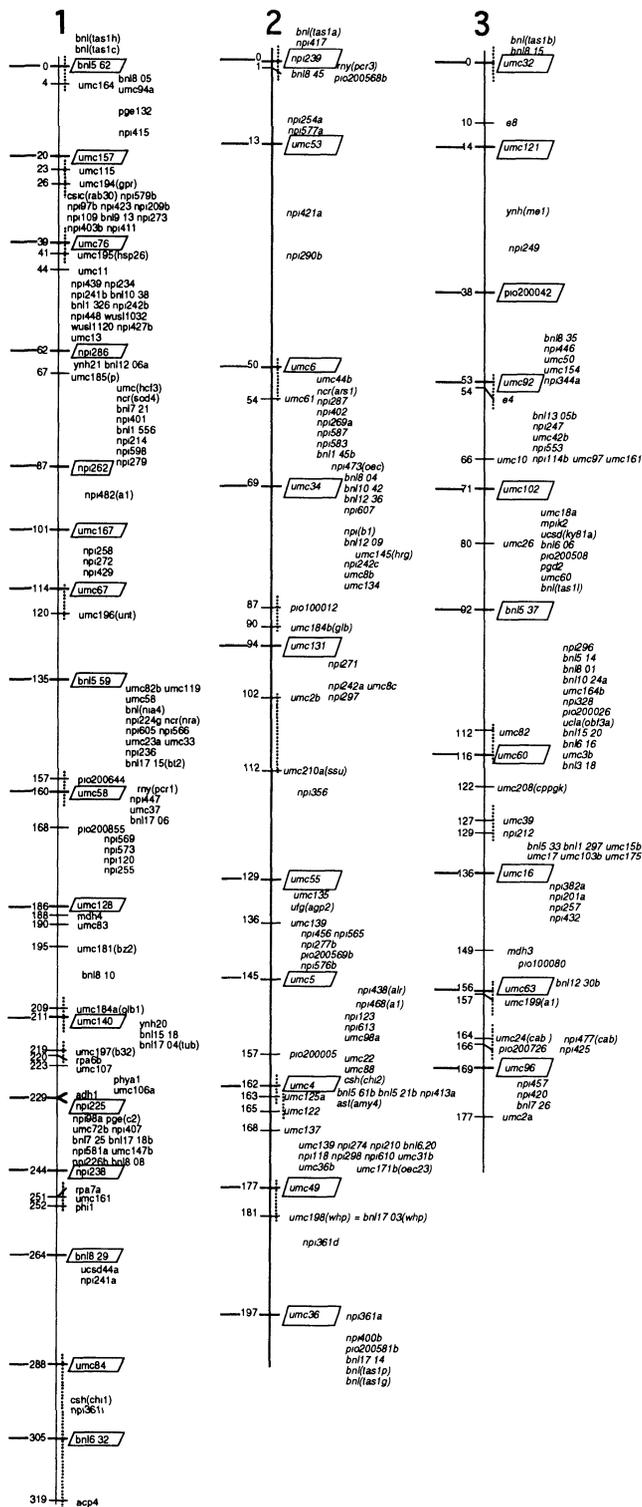


Fig. 1. RFLP map of molecular markers in maize.

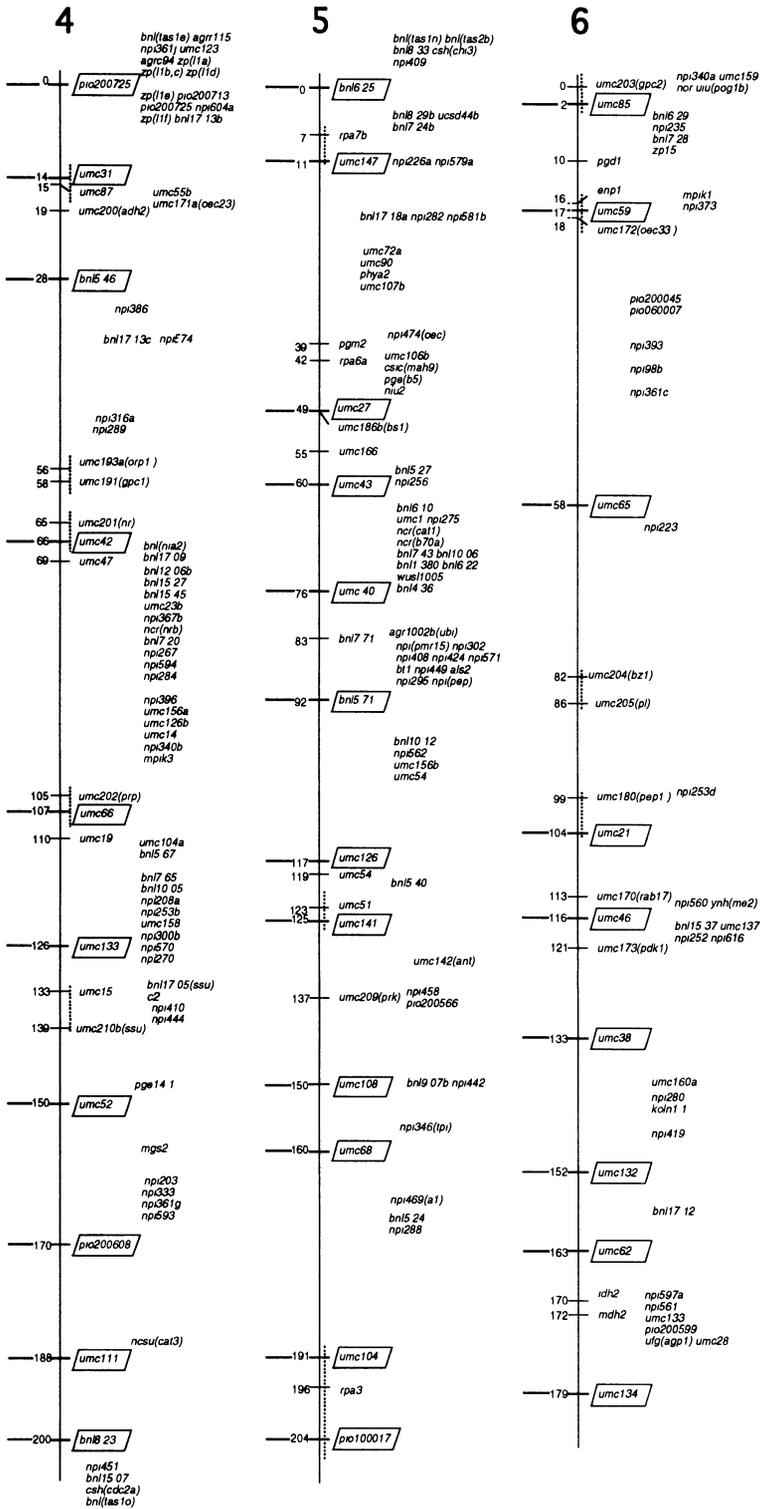


Fig. 1. Continued.

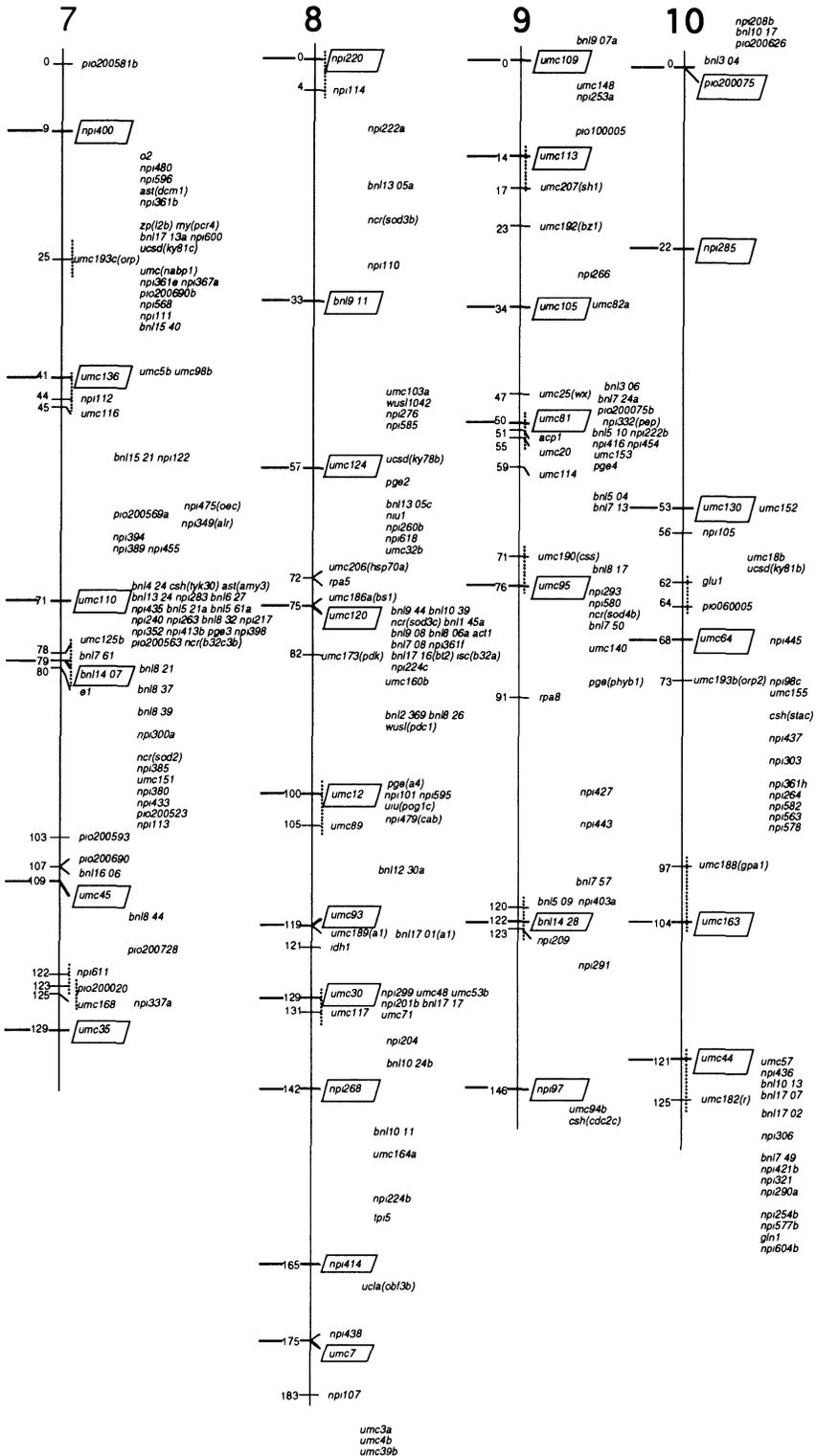


Fig. 1. Continued.

References

- Beavis, W D and Grant, D (1991) A linkage map based on information from four F₂ populations of maize (*Zea mays* L.) *Theor Appl Genet* 82 636-644
- Burr, B and Burr, F (1991) Recombinant inbreds for molecular mapping in maize theoretical and practical considerations *Trends Genet* 7 55-60
- Burr, B, Burr, F A, Thompson, K H, Albertsen, M C and Stuber, C W (1988) Gene mapping with recombinant inbreds in maize *Genetics* 118 519-526
- Gardiner, J, Coe, E H, Melia-Hancock, S, Hoisington, D A and Chao, S (1993) Development of a core RFLP map in maize using an immortalized F₂ population *Genetics* 134 917-930
- Helentjaris, T, Slocum, M, Wright, S, Schaefer, A and Nienhuis, J (1986a) Construction of genetic maps in maize and tomato using restriction fragment length polymorphisms *Theor Appl Genet* 72 761-769
- Helentjaris, T, Weber, D F and Wright, S (1986b) Use of monosomics to map cloned DNA fragments in maize *Proc Natl Acad Sci U S A* 83 6035-6039

14. RFLP map of peanut

TRACY HALWARD¹, H. THOMAS STALKER² and GARY KOCHERT¹

¹ Department of Botany, University of Georgia, Athens, GA 30602-7271, U.S.A.; ² Department of Crop Science, North Carolina State University, Raleigh, NC 27695-3281, U.S.A.

Contents

1. Introduction	246	5. Applications of the map to peanut breeding and genetics	255
2. Libraries and probe sources	249	6. References	258
3. Mapping population	250		
4. The map	251		

1. Introduction

Cultivated peanut (*Arachis hypogaea* L.) provides a significant source of oil and protein for large segments of the population, particularly in the less developed regions of Asia, Africa, and South America. In the United States, peanut is considered a high-value cash crop of regional importance, with production concentrated in the Southeast region of the country along with parts of Texas and Oklahoma. Domestically, peanuts are grown primarily for use in the snackfood, peanut butter and confection industries, but also serve as an excellent source of mono-unsaturated cooking oil, as well as a source of meal for livestock.

The genus *Arachis*, a member of the Leguminosae family, is native to South America, with central Brazil believed to be the center of origin. Wild species of *Arachis* are widely distributed from the Atlantic Ocean to the foothills of the Andes Mountains, and from the mouth of the Amazon in the north to Uruguay in the south (Stalker and Moss 1987). Peanuts have been found in such diverse environments as rock outcroppings, heavy to sandy soils, marshy areas and streams, forest-grassland margins, and from sea level to an elevation of 1600 meters (Valls et al. 1985). Although a complete taxonomic treatise has not been published to date, Gregory et al. (1973) divided the genus into seven sections based on morphology, geographic distribution, and cross compatibilities. While alternative taxonomic schemes have been proposed (Ressler 1980), the one outlined by Gregory et al. (1973) is the most widely used. Cultivated peanut, an allotetraploid ($2n = 4x = 40$), has been assigned to section *Arachis* which also

contains the proposed allotetraploid progenitor *A. monticola*, Krap. et Rig. In addition, at least 20–25 diploid species ($2n = 2x = 20$) have been described as belonging to section *Arachis*. Cultivated peanut has been further subdivided into two subspecies, each containing two botanical varieties: subsp. *hypogaea*, var. *hypogaea* (virginia); subsp. *hypogaea*, var. *hirsuta* (peruvian runner); subsp. *fastigiata*, var. *fastigiata* (valencia); and subsp. *fastigiata*, var. *vulgaris* (spanish type) (Krapovickas and Rigoni 1960). Cultivated peanut is thought to have originated in Bolivia at the base or in the foothills of the Andes mountains. This region is an important source of variability for the subspecies *hypogaea* and is the only region where *A. monticola* is known to occur (Krapovickas 1969).

Abundant germplasm resources of both the cultivated species and related wild species are available to peanut breeders. However, a large portion of these potential genetic resources have been inadequately evaluated for useful traits, and most peanut breeding programs have traditionally relied on the crossing of elite breeding lines for developing improved cultivars. As a result, the germplasm base of domesticated peanut is extremely narrow. Although considerable levels of morphological variability have been observed among the germplasm resources of cultivated peanut (Wynne and Halward 1989; ICRISAT 1982), very little genetic polymorphism has been detected with molecular markers within *A. hypogaea*. Grieshammer and Wynne (1990) evaluated a broad range of genotypes for isozyme variation and observed very little polymorphism within the cultivated germplasm. The researchers concluded that isozyme analysis would not be useful for characterizing genetic diversity in cultivated peanut. Using RFLPs, little variation has been observed among cultivars (Kochert et al. 1991) or exotic germplasm lines (Halward et al. 1991) of *A. hypogaea*. Similarly, studies using RAPD marker analysis have revealed little genetic polymorphism within *A. hypogaea* (Halward et al. 1992). Conversely, abundant polymorphism has been detected among related wild species of peanut in section *Arachis* with isozymes (Stalker et al. 1990), RFLPs (Kochert et al. 1991), and RAPD markers (Halward et al. 1992).

The apparent contradiction between the abundance of morphological variability observed and the lack of detectable genetic polymorphisms within the cultivated germplasm is not uncommon. Similar results have been observed in tomato (*Lycopersicon* sp.) (Miller and Tanksley 1990), melons (*Cucumis* sp.) (Shattuck-Eidens et al. 1990), soybean (*Glycine* sp.) (Keim et al. 1990), and common bean (*Phaseolus* sp.) (Gepts 1991). Morphological traits are often controlled by a few major genes and may be subjected to intense selection pressure. As a result, morphological variation is likely to increase during domestication. On the other hand, biochemical and molecular markers, which are not subject to direct selection, often decrease during domestication of a species (Gepts 1991). A review of genetic studies indicates that a large number of phenotypic traits in cultivated peanut are controlled by a few major genes with expression influenced by the action of modifier genes and epistatic interactions among loci (Wynne and Halward 1989). Intense selection for a few major genes affecting obvious morphological traits during domestication of *A.*

hypogaea which, after polyploidization, had been cut off from introgression and gene exchange with its wild ancestors could explain the lack of variability observed at the molecular level in cultivated peanut (D. Williams, pers. comm.). The highly self-pollinating nature of cultivated peanut would serve to enhance genetic isolation. A similar situation has been observed in domesticated bean (*Phaseolus vulgaris*) (Gepts 1991). Results from the isozyme, RFLP, and RAPD marker studies mentioned above support the hypothesis of Smartt and Stalker (1982) that *A. hypogaea* originated from a single polyploidization event. Thus, including exotic germplasm lines of *A. hypogaea* in a breeding program will result in the addition of only limited genetic variability due to the severe narrowing of the germplasm base that occurred during evolution of the cultivated species. The reduction in genetic diversity that accompanied the evolution of cultivated peanut represents a genetic bottleneck for peanut improvement.

In light of the relatively narrow germplasm base of cultivated peanut, a greater emphasis should be placed on the evaluation and utilization of related wild species to enhance the genetic variability available for the development of improved cultivars. The relative levels of variability observed at the molecular level for wild and cultivated peanut species is consistent with the variability observed for disease and insect resistance and for tolerance to a variety of environmental stresses. Extensive screening of wild *Arachis* species has revealed these potential genetic resources to be valuable as sources of disease and insect resistance (ICRISAT 1982; Moss 1980; Subrahmanyam et al. 1982), tolerance to environmental stresses (ICRISAT 1982), and variation for protein and oil quality (Cherry 1977; Amaya et al. 1977; Young et al. 1973). The exploitation of these valuable genetic resources would be greatly enhanced through the use of molecular markers to tag and follow the introgression of chromosome segments containing desirable traits from the wild species into cultivated peanut, and through the development of a genetic linkage map in peanut to expedite the location and transfer of these chromosomal regions.

Basic genetic research on peanut has not proceeded as rapidly as it has in many other species of agricultural importance. This is at least partially due to the limited acreage devoted to domestic peanut production, as compared to other major agronomic crops, and the relative importance of peanut as a staple crop only in less developed regions of the world. As a result, little information is known about the molecular biology or evolutionary history of the genus *Arachis*. Although there have been a few reports of linkage between various morphological traits (Badami 1928; Patel et al. 1936; Patil 1965; Coffelt and Hammons 1973; Stalker et al. 1979), there is no genetic linkage map available for peanut. However, a collaborative effort is currently underway to develop a linkage map based on a combination of molecular and conventional markers. The basic framework of the map, which is based on RFLPs, is being developed in our laboratory in the Department of Botany at the University of Georgia. The mapping populations are also being evaluated for a number of agronomic traits by researchers in the Department of Plant Pathology at the University of

Georgia and the Departments of Crop Science and Plant Pathology at North Carolina State University, with the goal of eventually producing a genetic map containing both conventional and molecular markers. The present status of the genetic map in peanut will be presented here.

2. Libraries and probe sources

Both random genomic peanut clones and cDNA clones have been used in the development of the linkage map. A genomic DNA library was constructed using the peanut cultivar 'Florunner' (*Arachis hypogaea*, subsp. *hypogaea*) as the genomic DNA source. The library was constructed by cloning gel-isolated *Pst*I fragments (1.0–2.0 kb in length) into pUC8 plasmids and transforming the recombinant plasmids into DH5 α strains of *E. coli*, according to the procedures of Sambrook et al. (1989). Recombinant clones were selected on IPTG-Xgal plates, and plasmids were isolated by a miniprep procedure (Wilimzig 1985). Recombinant plasmids were digested with *Pst*I and subjected to electrophoresis on 0.8% agarose gels. The molecular weight of each insert was determined by comparison to molecular weight standards, and blots were prepared on nylon filters (Gene Screen Plus, Du Pont) following the method of Southern (1975). Total peanut DNA was labeled with ³²P-dCTP by nick-translation (Rigby et al. 1977) and hybridized to the filters to detect inserts containing repeated sequences. Cotton chloroplast DNA (compliments of G. Galau, Dept. of Botany, Univ. of Georgia) was used to screen the library for members containing chloroplast DNA inserts. Those inserts that showed no signal with either chloroplast or total peanut DNA were assumed to represent low-copy number nuclear sequences and were selected for RFLP analysis.

Two separate cDNA libraries were prepared, one from root tissue and one from shoot tissue, using the Stratagene ZAP-cDNA GigapackII Gold Cloning Kit, according to the manufacturer's instructions. Approximately 200 seeds of the peanut cultivar 'GK-7' were germinated in moist paper towels to yield a total of 10–15 grams each of root and shoot tissue for the two libraries. Poly A⁺ RNA was isolated independently from young shoots and young roots according to the procedures described in Hong et al. (1989) and used to construct the two cDNA libraries.

Clones in the random genomic library are being maintained as glycerol cultures and as minipreparations of plasmids as described in Sambrook et al. (1989). The cDNA libraries are being maintained as whole phage for current use in addition to being cloned into Bluescript plasmids for long-term storage of clones. All clones from both the random genomic and cDNA libraries are available for distribution to any researchers who are interested in using them for genetic analyses. Genomic clones can be distributed as plasmid minipreps or as PCR products following amplification with M13 forward and reverse primers; clones from either cDNA library are available as small quantities of PCR product which can subsequently be amplified for use in random primer labeling

reactions using forward and reverse M13 primers and the following amplification conditions: 94 °C for 1 min. 20 sec. (denaturation); 37 °C for 2 min. (annealing); and 72 °C for 3 min. (extension) for 30 cycles.

3. Mapping population

A number of factors were taken into consideration when selecting an appropriate population for developing an RFLP map in peanut. Insufficient variability for RFLPs or RAPD markers exists within cultivated peanut to allow construction of a genetic linkage map directly in *A. hypogaea*. In addition, since cultivated peanut is an allotetraploid, segregation of traits is inherently complex. The genetic basis for inheritance of most agronomically important traits in peanut is not completely understood. Although most traits in cultivated peanut follow a diploid pattern of inheritance, many appear to be under the influence of duplicate loci (Wynne and Halward 1989). As a result, segregation analysis is more difficult in the allotetraploid cultivated species than in the related diploid species. A complete series of aneuploids, often used in the construction of RFLP maps (Beckmann and Soller 1986), is not available in peanut. For these reasons, the diploid species of *Arachis* were considered the best choice for a mapping population. The main advantage to constructing an RFLP map in diploid peanut is the abundance of RFLP variability observed among the wild species. The relative ease of analyzing segregation data in a diploid cross versus a cross between allotetraploids provides an additional advantage to mapping in wild *Arachis* species. RFLP maps that have been developed using populations derived from interspecific crosses between a cultivated species and a related wild species have proven useful for cultivar improvement in such crops as tomato (Miller and Tanksley 1990) and soybean (Keim et al. 1990). Although the peanut RFLP map was developed using populations derived from crosses among wild *Arachis* species it should also be useful for cultivar improvement programs, especially when applied to following the introgression of chromosome segments from wild species into cultivated peanut.

In peanut, we were fortunate to have available several F₂ populations derived from interspecific crosses between various diploid species of *Arachis*. Two of these populations were chosen for map construction. Simultaneous mapping in two independent populations has allowed us to evaluate the homosequentiality of chromosomes and to verify linkage relationships among populations. The two mapping populations we are evaluating [(*A. stenosperma* × *A. cardenasii*) and (*A. duranensis* × *A. diogeni*)] were both developed at North Carolina State University. A greater number of clones have revealed polymorphisms between *A. stenosperma* and *A. cardenasii* than between *A. duranensis* and *A. diogeni*. Therefore, we are concentrating mapping efforts in the F₂ population resulting from the *A. stenosperma* × *A. cardenasii* cross, and are evaluating the *A. duranensis* × *A. diogeni* F₂ population primarily for the

purpose of comparative mapping in different peanut populations and for verifying linkage relationships. A total of 87 F₂ individuals are being evaluated in the *A. stenosperma* × *A. cardenasii* cross, and 83 F₂ individuals are being evaluated in the *A. duranensis* × *A. diogenii* cross.

The mapping populations were not originally developed for the purpose of RFLP analysis, therefore our F₂ populations consist of bulked progeny from several F₁ plants. Given the highly self-pollinating nature of peanut, the original parents of the crosses should have been essentially homozygous and the F₁s from which our mapping populations were developed should have been genetically similar. The original crosses from which the populations were developed have been remade such that a single F₁ gave rise to each of the F₂ populations. The original F₁s are being maintained as cuttings and seeds of the new F₂s have been planted to produce new mapping populations. These new populations, derived from single F₁ plants, will be evaluated using the RFLP markers on the existing map in order to confirm the accuracy of the linkage groups developed in the original mapping populations. Once linkage relationships are confirmed we will begin mapping in the new F₂ populations. These new F₂ populations and the availability of the F₁s will also enable us to begin mapping RAPD markers with a higher degree of certainty. Both mapping populations will be available for distribution as cuttings of original plants to any researchers interested in utilizing the material in their programs.

4. The map

Genomic DNA was isolated from the parents, F₁s, and F₂s using a crude nuclear preparation as described in Kochert et al. (1991). The parents used to generate the two mapping populations were screened for clones that revealed polymorphisms between them in order to identify clones that would be segregating in the F₂ generation. Survey filters consisted of genomic DNA from each of the parents (*A. stenosperma*, *A. cardenasii*, *A. duranensis*, and *A. diogenii*) digested with seven restriction enzymes (*Bam*HI, *Dra*I, *Eco*RI, *Eco*RV, *Hae*III, *Hind*III, *Rsa*I). Several enzymes were used in the initial screening of the parents, as previous results (unpub. data) indicated that a given clone often revealed polymorphism with some enzymes but not others. By simultaneously screening the parents with a number of probe-enzyme combinations, a greater number of polymorphic clones could be detected. Clones that were found to reveal polymorphisms between the parents were used to evaluate the F₂ mapping populations for segregation. Initially, only those clones that revealed polymorphism at a single locus were used in construction of the RFLP linkage map. Once a framework was in place, clones detecting multiple loci were added to the developing map. A total of 100 random genomic clones and 300 cDNA clones have been evaluated for polymorphism to date. Of these, 15 (15%) of the genomic clones and 190 (63%) of the cDNA clones were polymorphic with one or more enzyme. Of those clones that were polymorphic, 7 (47%) genomic and

92 (48%) cDNA clones, respectively, detected a single polymorphic locus; while 8 (53%) and 98 (52%) detected polymorphisms at multiple loci.

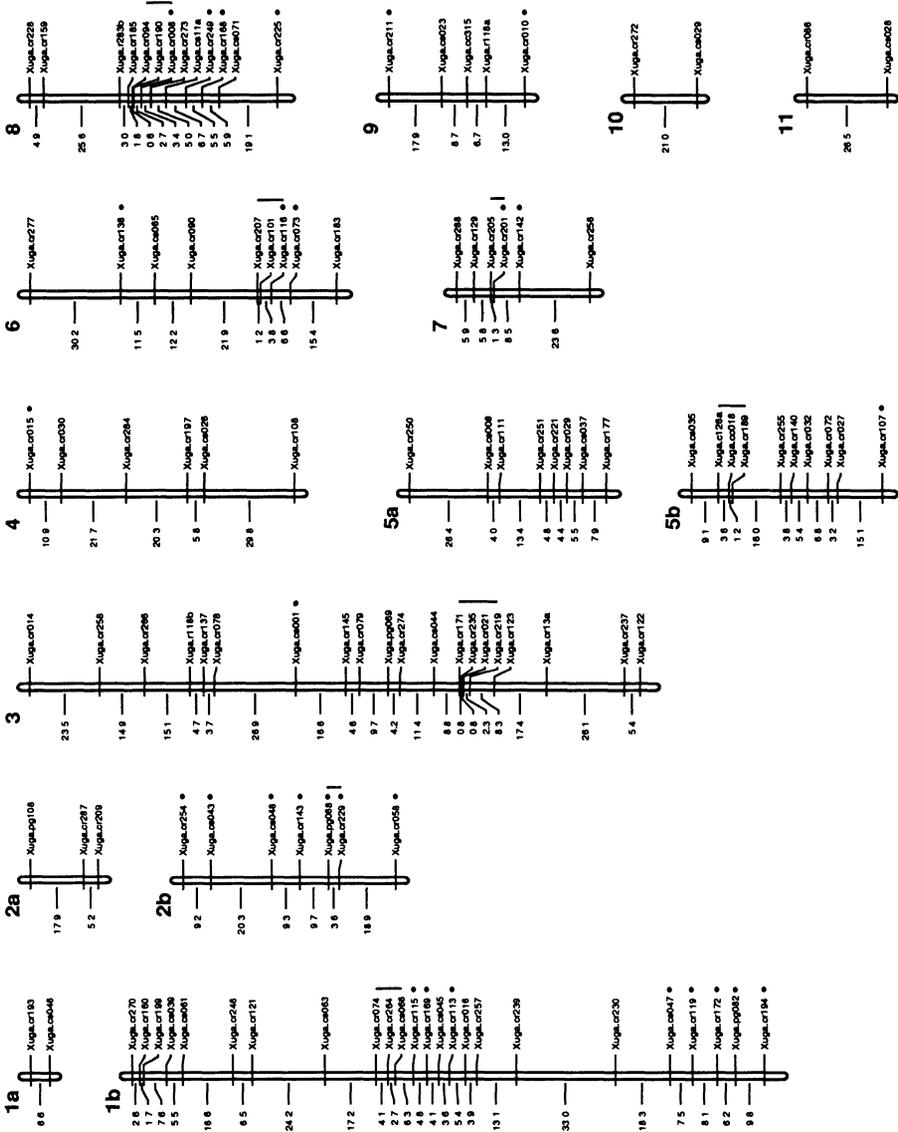
A large number of probes detected polymorphisms with more than one restriction enzyme, suggesting that many of the RFLP markers identified in peanut resulted from DNA rearrangements. Similar results have been observed in a number of other species including soybean (Apuya et al. 1988) and rice (McCouch et al. 1988). Of the cDNA clones evaluated, 38 produced ladders with one or more restriction enzyme suggesting the presence of a number of repeat sequences within the peanut genome. Unfortunately, no variability was observed among the parents of either mapping population for these clones and therefore they could not be mapped. Mapping these complex probes would allow us to determine whether the repeat sequences detected are scattered among linkage groups throughout the genome or are clustered within particular linkage groups. The information obtained from such analyses would provide insight into the evolution of the peanut genome. Therefore, we will continue to look for clones which produce ladders that reveal polymorphisms between the parents so that these can then be mapped in the future.

Several probes hybridized to more than one polymorphic locus. While the majority of these appeared to be cosegregating, a few were found to be segregating independently, indicating the presence of duplicated loci within the peanut genome. Unfortunately, the majority of the clones detecting duplicated loci produced very complex banding patterns which were difficult to score and could not be mapped with accuracy. We chose not to apply these markers to the map. As the map develops and becomes saturated with markers, it would be desirable to attempt to map some of these more complex probes in order to evaluate the extent to which duplicated loci have evolved within diploid *Arachis* species. Only two clones detecting duplication could be mapped with confidence; both of these clones hybridized to loci that were repeated on different linkage groups. The level of duplication observed within the peanut genome may have important implications regarding genome evolution in the genus. High levels of sequence duplication, similar to those detected in peanut, have been observed in *Brassica* sp. (Slocum et al. 1990; Song et al. 1991). These levels of duplication are much greater than those reported for tomato (Bernatsky and Tanksley 1986), potato (*Solanum* sp.) (Bonierbale et al. 1988), or rice (*Oryza* sp.) (McCouch et al. 1988).

To analyze segregation in the mapping population, the F₂ progeny were scored as either 'A' (homozygous like parent A), 'B' (homozygous like parent B), or 'H' (heterozygous). Chi-square analyses were conducted to determine goodness-of-fit of the segregation data to the expected 1:2:1 Mendelian ratio. Thirty three of the 103 mapped loci showed deviation from the expected ratio ($p = 0.05$). Most of these loci were found to have an excess of one or the other parental type, while four had an excess number of heterozygotes. The segregation data obtained was analyzed using the MAPMAKER computer package which is specifically designed for the construction of primary genetic linkage maps (Lander et al. 1987). The locus arrangements and map distances

for each linkage group were determined based on the output from the MAPMAKER program using the Kosambi mapping function and the constraints of minimum LOD score of 3.0 and a maximum recombination frequency (θ) of 0.25. RFLP loci detected by different probes were assigned different numbers using the following convention: Xuga.pg (or cr, or cs) -#, where pg indicates a genomic clone and cr and cs indicate cDNA clones from the root and shoot libraries, respectively. Multiple segregating loci detected by a single probe were assigned the same number followed by a letter (a, b, etc.) to indicate each duplicated locus. In addition to random genomic and cDNA clones, one cDNA clone (Xuga.cc 315 obtained from Dr. A. Abbott, Clemson University) representing the gene for stearoyl ACP desaturase has been mapped in peanut. The segregating loci are presently distributed among 12 linkage groups (Fig. 1). However, two of the linkage groups contain only two markers each and are expected to link up with one of the larger linkage blocks as additional markers are added to the map. Eventually, the map should consist of 10 linkage groups corresponding to the haploid chromosome number of diploid *Arachis* sp. To date, a total map distance of 1400 cM has been covered to 20 cM resolution. This is estimated to represent approximately 80% coverage of the peanut genome.

Both conventional RFLP markers and RAPD markers can be useful for expanding the existing peanut map, each having advantages and disadvantages as molecular markers. The main advantages to using RAPD markers are the speed with which large numbers of markers can be added to the map, the relatively small amounts of DNA required for analysis, and the elimination of the need for radioisotopes, restriction enzymes, and the laborious procedure of Southern blotting. Disadvantages of RAPD markers include the production of complex banding patterns with most primers, making comparisons of mapped markers among populations or laboratories difficult; the degree of reproducibility among different DNA extraction preparations and different researchers is still being debated; and the dominant-recessive (presence or absence of bands) nature of the majority of RAPD markers restricts the information that can be derived from linkage analysis. Conversely, RFLPs behave as codominant markers giving the maximum amount of information obtainable from a mapping population. Other advantages of RFLPs as molecular markers are the high degree of reproducibility, and the relatively large number of simple banding patterns produced for most organisms evaluated, particularly if libraries have been screened for single copy clones. However, the larger amounts of DNA required for RFLP analysis relative to RAPD marker analysis, and the use of radioisotopes, large quantities of restriction enzymes, and Southern blotting procedures for evaluating RFLPs are all disadvantages associated with using RFLPs as molecular markers. One approach to mapping in peanut, in order to expand the map as quickly as possible, would be to use RFLPs to construct a solid framework for the map and to subsequently fill it in with RAPD markers. Techniques developed recently for selecting markers in specific chromosomal regions using RAPD markers and



pooled DNA samples from segregating populations (Michelmore et al. 1991; Giovannoni et al. 1991) might also be useful for obtaining additional markers in specific areas of interest, such as in the bracketing of QTLs. However, such approaches will not be effective in peanut until a more extensive basic mapping framework is in place.

We are also investigating the use of microsatellite markers for mapping in peanut. Southern blots of peanut DNA digested with restriction enzymes were probed with a number of synthetic oligomers composed of simple nucleotide repeats such as (GT) n or (CAC) n . The initial survey indicated that such microsatellites are present in peanut. When genomic DNA from a diverse group of exotic germplasm lines of *A. hypogaea* was amplified using PCR with microsatellite primers isolated from rice, slight differences in banding patterns were observed among the genotypes evaluated. The random genomic peanut library is being probed with a number of short oligomers to isolate clones containing microsatellites. After sequencing the region found to contain a microsatellite, PCR primers flanking the microsatellite can then be used to convert the markers to sequence tagged sites (STSs) as was successfully done in rice (*Oryza sativa*) (Zhao and Kochert 1992). Microsatellites have potential advantages as molecular markers in that they can be analyzed with simple procedures utilizing PCR amplification with flanking markers, and because they are often more polymorphic than conventional RFLPs or RAPD markers (Edwards et al. 1991; Nanda et al. 1991; Stallings et al. 1991). This latter advantage will be especially important in transferring the presently developing peanut map to cultivated peanut since no other molecular markers evaluated to date have revealed sufficient variability within *A. hypogaea* for mapping applications.

5. Applications of the map to peanut breeding and genetics

Peanut chromosomes are small with few distinctive cytogenetic markers, making identification of individual chromosomes quite tedious (Stalker and Dalmacio 1986). In addition, no classical genetic map exists for peanut, and no trait has yet been mapped to a specific peanut chromosome (Stalker 1991). Without the availability of a genetic map, it has not been possible to utilize molecular markers in peanut breeding, or to combine molecular and conventional genetic techniques in peanut improvement programs. This is most

←

Fig 1 A peanut RFLP map developed using an F₂ population derived from the cross *Arachis stenosperma* × *A. cardenasu*. Loci were ordered using the MAPMAKER program (minimum LOD 3.0, maximum recombination 0.25). RFLP markers were either random genomic clones or cDNA clones. Random genomic clones have 'pg' in the clone designation, cDNAs from a root library have 'cr' and cDNAs from a shoot library have 'cs'. Areas of closely linked markers where the order is uncertain are indicated by a vertical bar next to the affected markers. Markers which segregated with ratios significantly different from the expected 1:2:1 are denoted by a dot (•)

unfortunate, as molecular markers provide powerful discriminatory tools to the plant breeder allowing for increased efficiency of conventional breeding programs (Beckmann and Soller 1986; Tanksley et al. 1989).

As the peanut RFLP map continues to develop, a number of potential advantages will be realized in peanut improvement programs. Our understanding of genetic segregation analysis and linkage relationships will be greatly enhanced as molecular markers accumulate in peanut. The chromosomal location and distribution of duplicated regions within the peanut genome will allow for a better understanding of the underlying basis for the complex patterns of segregation often observed in peanut, particularly among the progeny of intersubspecific crosses (Wynne and Coffelt 1980; Wynne and Halward 1989).

Given the relatively narrow germplasm base available in *A. hypogaea*, and the abundance of polymorphism observed for both morphological traits and molecular markers in wild species of *Arachis*, increased emphasis should be placed on the utilization of wild relatives in peanut improvement programs. This is perhaps the area for which a comprehensive RFLP map will be of most benefit to peanut breeders. A number of methods have been investigated for their potential utilization in the introgression of traits from related diploid wild species into cultivated peanut (Simpson 1991; Stalker 1992). One such method, the 'hexaploid route', involves making crosses between diploid wild species and the tetraploid cultivated species to produce triploid hybrids. These sterile hybrids are then doubled with colchicine to form hexaploids, which are subsequently backcrossed to cultivated peanut producing pentaploid hybrids. Repeated selfing of these pentaploids often results in the spontaneous loss of chromosomes with eventual stabilization of progeny at the tetraploid level (Stalker 1992). In addition to investigations involving the above route of introgression, both amphidiploids and autotetraploids have been constructed among several of the wild diploid species, and these have then been crossed directly with *A. hypogaea* (Singh 1986a,b). Only limited success has been achieved using any of the above methods of introgression as part of a conventional peanut breeding program. In early generations, the progeny resulting from interspecific hybridizations between wild and cultivated peanut species usually appear very different from *A. hypogaea* and retain many of the undesirable traits of the wild parent(s). Subsequent selection for agronomic traits during selfing generations typically results in the loss of desirable characters the breeder was attempting to introgress from the wild parent. The net result is an enormous waste of time and resources. The development of a genetic linkage map in peanut will greatly enhance the ability of breeders to tag and follow the introgression of specific chromosome segments linked to desirable traits from wild species into breeding lines of cultivated peanut. The developing peanut RFLP map will be useful for monitoring introgression following the mapping of valuable agronomic traits in wild species.

An example of the application of introgression analysis to peanut can be illustrated using an interspecific population which was derived from a cross

between *A. hypogaea* × *A. cardenasii* that contains highly variable hybrid derivatives which are stable at the tetraploid level (Stalker et al. 1979). Many of the lines appear to have traits that were introgressed from *A. cardenasii* and which are not found in the cultivated parent, including large seed size, high yields, and resistance to rust (*Puccinia arachidis* Speg.), early and late leafspot [*Cercospora arachidicola* Hori and *Cercosporidium personatum* (Berk. and Curt.) Deighton, respectively], nematodes, and a number of insect pests (Stalker 1991). Breeding lines from this population could be analyzed for molecular markers indicating the presence of specific chromosome segments introgressed from *A. cardenasii*, and to correlate the presence of such segments with introgressed characters. It will then be possible to estimate the number of introgressed segments as well as their chromosomal location(s). The ability to identify and follow specific chromosome segments from wild *Arachis* species through three generations of backcrossing following interspecific hybridization with cultivated peanut has already been demonstrated (unpub. data). In addition, Kochert, Halward and Stalker (in prep.) have proposed a breeding scheme involving the use of molecular marker-assisted interspecific hybridization aimed at the development of multiple populations, each with a number of desirable characters, for general use in peanut breeding programs. The proposed method would involve the development of several breeding populations derived from interspecific hybridizations among various wild *Arachis* species. The F₁ interspecific hybrids would then be backcrossed to cultivated peanut. Plants in the BC₂ or BC₃ generation would be selected and families would subsequently be developed by a combination of selfing and single seed descent for several generations to produce lines that are homozygous for specific introgressed segments which could be detected by molecular marker analysis. Introgressed segments would then be correlated with specific agronomic traits and eventually introgressed into cultivated breeding lines. By providing a means for breeders to screen interspecific populations in the seedling stage, rather than evaluating large segregating populations in the field for the trait of interest, molecular markers will become a valuable tool for enhancing the efficiency of peanut breeding programs involving the introgression of genes from related wild species.

Protocols for transformation in peanut are being investigated in a number of laboratories (Weissinger 1991; W. Parrott, pers. comm.; P. Ozias-Akins, pers. comm.). From these investigations, it appears that a satisfactory system for peanut transformation and regeneration will be available in the near future. Once protocols for peanut transformation and regeneration are in place, these can be combined with the use of map-based cloning procedures, which are being rapidly developed and should become widely applicable in the future (Young 1990; Collins 1991). As these procedures become routinely available, knowing the map location of desirable genes will greatly enhance their utilization in peanut breeding programs, making the peanut RFLP map a valuable resource for use in cultivar improvement programs.

6. References

- Amaya, F, Young, C T, Hammons, R O and Martin, G (1977) The tryptophan content of the U S commercial and some South American wild genotypes of the genus *Arachis* a survey *Oleagineux* 32 225-229
- Apuya, N, Frazier, B L, Keim, P, Roth, E J and Lark, K G (1988) Restriction length polymorphisms as genetic markers in soybean, *Glycine max* (L) Merr *Theor Appl Genet* 75 889-901
- Badami, V K (1928) *Arachis hypogaea* (the groundnut) Ph D thesis, Cambridge England University Library Inheritance Studies
- Beckmann, J S and Soller, M (1986) Restriction fragment length polymorphisms and genetic improvement of agricultural species *Euphytica* 35 111-124
- Bernatzky, R and Tanksley, S D (1986) Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences *Genetics* 112 887-898
- Bonierbale, M W, Plaisted, R L and Tanksley, S D (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato *Genetics* 120 1095-1103
- Cherry, J P (1977) Potential sources of peanut seed proteins and oil in the genus *Arachis* *J Agric Food Chem* 25 186-193
- Coffelt, T A and Hammons, R O (1973) Influence of sizing peanut seed on two phenotypic ratios *J Hered* 64 39-42
- Collins, F S (1991) Of needles and haystacks - finding human disease genes by positional cloning *Clin Res* 39 615-623
- Edwards, A, Civitello, A, Hammond, H A and Caskey, C T (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats *Am J Hum Genet* 49 746-756
- Gepts, P (1991) Biotechnology sheds light on bean domestication in Latin America *Diversity* 7 49-50
- Giovannoni, J J, Wing, R A, Ganai, M W and Tanksley, S D (1991) Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations *Nucl Acids Res* 19 6553-6558
- Gregory, W C, Gregory, M P, Krapovickas, A, Smith, B W and Yarbrough J A (1973) Structure and genetic resources of peanuts. In *Peanuts - Culture and Uses*, pp 47-133 *Am Peanut Res Educ Assoc*, Stillwater, OK
- Griehammer, U and Wynne, J C (1990) Mendelian and non-Mendelian inheritance of three isozymes in peanut (*Arachis hypogaea* L) *Peanut Sci* 17 101-105
- Halward, T M, Stalker, H T, LaRue, E A and Kochert, G (1991) Genetic variation detectable with molecular markers among unadapted germplasm resources of cultivated peanut and related wild species *Genome* 34 1013-1020
- Halward, T M, Stalker, H T, LaRue, E A and Kochert, G (1992) Use of single-primer DNA amplifications in genetic studies of peanut *Plant Mol Biol* 18(2) 315-325
- Hong, J C, Nagao, R T and Key, J L (1987) Characterization and sequence analysis of a developmentally regulated putative cell wall protein gene isolated from soybean *J Biol Chem* 17 8367-8376
- ICRISAT (1982) Annual Report Patancheru, A P India
- Keim, P, Dier, B W, Olson, T C and Shoemaker, R C (1990) RFLP mapping in soybean association between marker loci and variation in quantitative traits *Genetics* 126 735-742
- Kochert, G, Halward, T M, Branch, W D and Simpson, C E (1991) RFLP variability in peanut cultivars and wild species *Theor Appl Genet* 81 565-570
- Krapovickas, A (1969) Evolution of the genus *Arachis* *Seminario Avanzado de Genetica Agricola para America Latina*, Maracay, Venezuela SAGA/B(d) 1-4
- Krapovickas, A and Rigoni, V A (1960) La nomenclatura de las subespecies y variedades de *Arachis hypogaea* L *Revista de Investigaciones Agricolas* 14 197-228
- Lander, E S, Green, P, Abrahamson, J, Barlow, A, Daly, M J, Lincoln, S E and Newburg, L (1987) MAPMAKER an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations *Genomics* 1 174-181

- McCouch, S R , Kochert, G , Yu, Z G , Wang, Z Y , Khush, G S , Coffman, W R and Tanksley, S D (1988) Molecular mapping of rice chromosomes Theor Appl Genet 76 815–829
- Michelmore, R W , Paran, I and Kesseli, R V (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis – a rapid method to detect markers in specific genomic regions by using segregating populations Proc Natl Acad Sci U S A 88 9828–9832
- Miller, J C and Tanksley, S D (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon* Theor Appl Genet 80 437–448
- Moss, J P (1980) Wild species in the improvement of groundnuts In R J Summerfield and A H Bunting (eds), Advances in Legume Science, Vol 1 Royal Botanic Gardens, Kew
- Nanda, I , Zischler, H , Epplen, C , Guttenbach, M and Schmid, M (1991) Chromosomal organization of simple repeated DNA sequences used for DNA fingerprinting Electrophoresis 12 193–203
- Patel, J S , John, C M and Seshadri, C R (1936) The inheritance of characters in the groundnut Proc Indian Acad Sci 3 214–233
- Patil, V H (1965) Genetic studies in groundnut (*Arachis hypogaea* L.) M Sc thesis, Poona University, India
- Ressler, P M (1980) A review of the nomenclature of the genus *Arachis* Euphytica 29 813–817
- Rigby, P , Dieckmann, M , Rhodes, C and Berg, P (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick-translation with DNA polymerase I J Mol Biol 113 237–251
- Sambrook, J , Fritsch, E F and Maniatis, T (1989) Molecular Cloning A Laboratory Manual Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Shattuck-Eidens, D M , Bell, R N , Neuhausen, S L and Helentjaris, T (1990) DNA sequence variation within maize and melon observations from polymerase chain reaction amplification and direct sequencing. Genetics 126 207–217
- Simpson, C E (1991) Global collaborations find and conserve the irreplaceable genetic resources of wild peanut in South America Diversity 7 59–61
- Singh, A K (1986a) Utilization of wild relatives in the genetic improvement of *Arachis hypogaea* L 7 Autotetraploid production and prospects in interspecific breeding Theor Appl Genet 72 164–169
- Singh, A K (1986b) Utilization of wild relatives in the genetic improvement of *Arachis hypogaea* L 8 Synthetic amphidiploids and their importance in interspecific breeding Theor Appl Genet 72 433–439
- Slocum, M K , Figdore, S S , Kennard, W C , Suzuki, J Y and Osborn, T C (1990) Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea* Theor Appl Genet 80 57–64
- Smartt, J and Stalker, H T (1982) Speciation and cytogenetics in *Arachis* In H E Patee and C T Young (eds), Peanut Science and Technology, pp 21–49 Am Peanut Res Educ Soc , Yoakum, TX
- Song, K M , Suzuki, J Y , Slocum, M K , Williams, P H and Osborn, T C (1991) A linkage map of *Brassica rapa* (syn *campestris*) based on restriction fragment length polymorphism loci Theor Appl Genet 82 296–304
- Southern, E M (1975) Detection of specific enzyme sequences among DNA fragments separated by gel electrophoresis J Mol Biol 98 503–517
- Stalker, H T (1991) A morphological appraisal of wild species in section *Arachis* of peanuts Peanut Sci 17 117–122
- Stalker, H T (1992) Utilizing *Arachis* germplasm resources In S N Nigam (ed), Groundnut – a Global Perspective, pp 281–295 ICRISAT, Patancheru, A P , India
- Stalker, H T and Dalmacio, R D (1986) Karyotype analysis and relationships among varieties of *Arachis hypogaea* L Cytologia 51 617–629
- Stalker, H T and Moss, J P (1987) Speciation, cytogenetics, and utilization of *Arachis* species Adv Agron 41 1–40
- Stalker, H T , Wynne, J C and Company, M (1979) Variation in progenies of an *Arachis hypogaea* × diploid wild species hybrid Euphytica 28 675–684

- Stalker, H T , Jones, T M and Murphy, J P (1990) Isozyme variability among *Arachis* species Proc Am Peanut Res Educ Soc 22 50
- Stallings, R L , Ford, A F , Nelson, D , Torney, D C , Hildebrand, C E and Moyzis, R K (1991) Evolution and distribution of (GT)_n repetitive sequences in mammalian genomes Genomics 10 807-815
- Subrahmanyam, P , McDonald, D , Gibbons, R W , Nigam, S N and Nevill, D J (1982) Resistance to rust and late leafspot diseases in some genotypes of *Arachis hypogaea* Peanut Sci 9 9-14
- Tanksley, S D , Young, N D , Paterson, A H and Bonierbale, M W (1989) RFLP mapping in plant breeding new tools for an old science Bio/Technology 7 257-264
- Valls, J F M , Rao, V R , Simpson, C E and Krapovickas, A (1985) Current status of collection and conservation of South American groundnut germplasm with emphasis on wild species of *Arachis* In Proc Int Workshop Cytogenet of *Arachis*, pp 15-33 ICRISAT, Patancheru, A P India
- Weissinger, A K (1991) Biotechnology for improvement of peanut (*Arachis hypogaea* L.) In Proc Int Workshop Cytogenet Arachis, ICRISAT, Patancheru, A P India
- Wilmzig, R (1985) LiCl method for plasmid minipreps Trends Genet 1 158
- Wynne, J C and Coffelt, T A (1982) Genetics of *Arachis hypogaea* In H E Patee and C T Young (eds), Peanut Science and Technology, pp 50-94 Am Peanut Res Educ Soc , Yoakum, TX
- Wynne, J C and Halward, T M (1989) Cytogenetics and genetics of *Arachis* Crit Rev Plant Sci 8(3) 189-220
- Young, C T , Waller, G R and Hammons, R O (1973) Variations in total amino acid content of peanut meal J Am Oil Chem Soc 50 521-523
- Young, N D (1990) Potential applications of map-based cloning to plant pathology Physiol Mol Plant Pathol 37 81-94
- Zhao, X P and Kochert, G (1992) Characterization and genetic mapping of a short, highly repeated, interspersed DNA sequence from rice (*Oryza sativa*) Mol Gen Genet 231(3) 353-359

15. *Phaseolus vulgaris*: the common bean

C. EDUARDO VALLEJOS

Department of Horticultural Sciences & Graduate Program in Plant Molecular and Cellular Biology, 1255 Fyfield Hall, University of Florida, Gainesville, FL 32611, U S A

Contents

1	Introduction	261	3	Some applications of RFLP markers	
2	Construction of the linkage map	262		in common bean	266
			4	References	269

1. Introduction

Gregor Mendel (1866) conducted the first genetic analysis of common beans. Mendel studied the inheritance of growth habit, and pod color and shape in a progeny between *P. vulgaris* and *P. nanus* (= *P. vulgaris*, bush type) in order to confirm his findings with peas. Unfortunately, further studies on the inheritance of flower and seed coat color were hampered by his use of interspecific hybrids between *P. nanus* and *P. multiflorus* (= *P. coccineus*), which are now known to yield aberrant ratios. Later, Shaw and Norton (1918) used intraspecific crosses and determined that pigmentation and pigmentation patterns of the seed coat are controlled by multiple independent factors. A few years later Sax (1923) began to identify the multiple components that determine the inheritance of these traits. A single factor was identified as responsible for pigmentation, while two linked factors were identified to control mottling; this appears to be the first report of linkage in beans. Furthermore, Sax (1923) was the first to report linkage between a Mendelian character (seed coat pigmentation) and a QTL (for seed size). Although the common bean was used as experimental material at the inception of genetics, its genetic characterization has lagged behind that of many other crop species.

The common bean is a diploid organism ($n = 11$) with relatively small chromosomes (Zheng et al. 1991) and a small genome estimated by flow cytometry to be 637 Mbp or 0.66 pg/1C (Arumuganathan and Earle 1991). It has also been estimated, via DNA reassociation kinetics, that 60% of the genome is comprised of single copy sequences (Talbot et al. 1984). The chromosome number and genome size of *P. vulgaris* are very similar to those of *P. acutifolius*

and *P. coccineus* (Arumuganathan and Earle 1991), both of which are partially compatible with the common bean and represent an important source of germplasm for plant improvement (Hucl and Scoles 1985). A rudimentary linkage map has been developed through the years with mostly morphological markers and a few isozymes (Bassett 1991; Vallejos and Chase 1991a,b). Electrophoretic analysis of the major seed storage protein (phaseolin) and a group of isozymes has led to the identification of a Mesoamerican and an Andean gene pool (Gepts et al. 1986; Koenig and Gepts 1989). Moreover, preliminary survey of DNA restriction fragment length polymorphisms showed that DNA probes can be used to differentiate the two groups because low levels of polymorphism were detected within each gene pool, but moderate levels were found between the gene pools (Chase et al. 1991).

2. Construction of the linkage map

Recently, a linkage map, based mainly on RFLP markers, has been constructed using a backcross progeny between a Mesoamerican breeding line and an Andean cultivar (Vallejos et al. 1992). A genomic library of size-selected (500–4000 bp) *Pst*I fragments was the main source of probes (Chase et al. 1991). This library was enriched for single copy fragments as 95% of the 362 clones tested yielded hybridization patterns typical of single copy sequences. Sixty percent of the clones tested revealed polymorphisms between the parental genotypes ('XR-235-1-1' and 'Calima') with at least one of four restriction enzymes (*Dra*I, *Eco*RI, *Eco*RV, and *Hind*III). More recently, 28 of the clones that had not detected RFLPs with any of the previously used enzymes were tested with four new enzymes: *Bam*HI, *Bg*III, *Kpn*I, and *Xba*I. About 50% of these clones have revealed polymorphisms with at least one of these enzymes. These results bring the estimated polymorphism between the parental genotypes at 80% with at least one of the 8 restriction enzymes. These enzymes differed in their ability to detect polymorphisms between the parental genotypes: *Kpn*I (40%); *Dra*I (42%); *Bam*HI, *Bg*III and *Hind*III (53%); *Eco*RI (62%); and *Eco*RV (64%).

Segregation data obtained from a backcross between the Mesoamerican breeding line 'XR-235-1-1' and the Andean cultivar 'Calima' have been used to assemble a linkage map for the common bean (Vallejos et al. 1992) with the aid of the computer software *Mapmaker* (Lander et al. 1987). This map (Fig. 1) was constructed by first establishing a framework order using stringent 'linkage criteria' (LOD 3.0, 25 cM, and three point LOD exclusion threshold – 3.0). Additional loci have been added later using either the *try* or *place* commands. The linkage between *Bng205a* and *Bng7* in group *F* has a LOD score of 2.22; although weak, this linkage is supported by data from two different progenies (Vallejos et al. 1992). The current map comprises 227 RFLP loci. Eight out of the 219 genomic clones listed in Table 1 hybridize to homologous sequences located in different linkage groups. The map also includes two loci of the chlorophyll a/b binding protein in groups *B* and *D*; these loci were identified

with pMB123, a mungbean cDNA clone (Thompson et al. 1983). Also included are: one phenotypically identified pigmentation locus (*P*), nine isozymes and nine seed proteins (Table 2). The seed proteins are: phaseolin, the α -amylase inhibitor proteins (identified by western blots with antibody provided by M. Chrispeels (Moreno and Chrispeels 1989), and other globulins (Vallejos and Chase 1991b). Electrophoretic variation at the protein level could be due to either variation at the DNA level and/or variation in post-translational events – protease processing and/or glycosylation. Nevertheless, Southern analysis of genomic blots with a phaseolin clone (Sun et al. 1981) has shown perfect co-segregation between a restriction fragment identified by this clone and phaseolin protein bands identified by SDS-PAGE. Thus, these results strongly suggest that variation at the protein level in phaseolin is due to variation at the DNA level and that the locus mapped with protein data corresponds to the structural gene of phaseolin. We have recently begun to use RAPD markers (Williams et al. 1990) and have added two of these markers to the map: *OAla* in *D* and *OAl0a* in *F* (Random primers were obtained from Operon Technologies, Inc; Alameda, CA) (Z.H. Yu and C.E. Vallejos, unpub.). In summary, a linkage map of the common bean has been assembled that includes 250 markers assigned to 150 loci/locus clusters and comprise 980 cM, or approximately 82% of the bean genome. The development of the bean genomic clones at the University of Florida was financed in part by a grant from the Agency for International Development (AID). These clones have been transferred to the International Center for Tropical Agriculture (CIAT) in Cali, Colombia, and can be requested from its Biotechnology Research Unit.

Table 1. List of loci identified with genomic clones (pBngN), their corresponding linkage group association, and relative position. The insert size of each clone is also included.

Locus	Linkage group	Relative position	Insert size (bp)	Locus	Linkage group	Relative position	Insert size (bp)
<i>Bng1</i>	<i>J</i>	3	1790	<i>Bng19</i>	<i>K</i>	12	2370
<i>Bng2</i>	<i>K</i>	1	1800	<i>Bng20</i>	<i>K</i>	7	1430
<i>Bng3</i>	<i>C</i>	12	1200	<i>Bng21</i>	<i>C</i>	2	1980
<i>Bng4</i>	<i>H</i>	7	1270	<i>Bng22</i>	<i>F</i>	7	1780
<i>Bng5</i>	<i>K</i>	12	1470	<i>Bng23</i>	<i>A</i>	8	620
<i>Bng6</i>	<i>K</i>	3	1400	<i>Bng24</i>	<i>K</i>	7	1070
<i>Bng7</i>	<i>F</i>	9	2800	<i>Bng25</i>	<i>J</i>	7	1100
<i>Bng8</i>	<i>G</i>	8	800	<i>Bng26</i>	<i>G</i>	13	1960
<i>Bng9</i>	<i>G</i>	8	4510	<i>Bng27</i>	<i>G</i>	6	1260
<i>Bng10</i>	<i>K</i>	13	4190	<i>Bng28</i>	<i>A</i>	4	2110
<i>Bng11</i>	<i>D</i>	19	3180	<i>Bng29</i>	<i>C</i>	9	1120
<i>Bng12</i>	<i>C</i>	17	2700	<i>Bng30</i>	<i>H</i>	4	1350
<i>Bng13</i>	<i>B</i>	5	1560	<i>Bng31</i>	<i>F</i>	5	1130
<i>Bng14</i>	<i>K</i>	9	3370	<i>Bng32</i>	<i>C</i>	4	950
<i>Bng15</i>	<i>A</i>	6	3220	<i>Bng33</i>	<i>C</i>	5	1100
<i>Bng16</i>	<i>C</i>	16	2040	<i>Bng34</i>	<i>D</i>	11	4330
<i>Bng17</i>	<i>D</i>	2	1600	<i>Bng35</i>	<i>K</i>	16	1020
<i>Bng18</i>	<i>F</i>	10	1100	<i>Bng36</i>	<i>H</i>	10	2500

Table 1. Continued.

Locus	Linkage group	Relative position	Insert size (bp)	Locus	Linkage group	Relative position	Insert size (bp)
<i>Bng37</i>	<i>K</i>	6	1250	<i>Bng87</i>	<i>G</i>	3	3350
<i>Bng38</i>	<i>F</i>	3	1800	<i>Bng88</i>	<i>G</i>	9	2510
<i>Bng39</i>	<i>H</i>	9	2300	<i>Bng89</i>	<i>D</i>	12	840
<i>Bng40</i>	<i>A</i>	21	800	<i>Bng90</i>	<i>D</i>	13	1850
<i>Bng41</i>	<i>H</i>	11	2780	<i>Bng91</i>	<i>J</i>	8	1470
<i>Bng42</i>	<i>A</i>	18	1500	<i>Bng92</i>	<i>J</i>	10	1130
<i>Bng43</i>	<i>F</i>	2	1060	<i>Bng93</i>	<i>K</i>	13	800
<i>Bng44</i>	<i>C</i>	5	970	<i>Bng94</i>	<i>G</i>	4	1460
<i>Bng45</i>	<i>D</i>	1	890	<i>Bng95</i>	<i>G</i>	10	1580
<i>Bng46</i>	<i>G</i>	9	2370	<i>Bng96</i>	<i>F</i>	6	1140
<i>Bng47</i>	<i>A</i>	2	2480	<i>Bng97</i>	<i>J</i>	8	1380
<i>Bng48</i>	<i>H</i>	11	1790	<i>Bng98</i>	<i>D</i>	17	1750
<i>Bng49</i>	<i>E</i>	4	1320	<i>Bng99</i>	<i>K</i>	12	1260
<i>Bng50</i>	<i>A</i>	3	850	<i>Bng100</i>	<i>I</i>	5	2160
<i>Bng51</i>	<i>K</i>	10	1150	<i>Bng101</i>	<i>J</i>	6	670
<i>Bng52</i>	<i>K</i>	4	1140	<i>Bng102</i>	<i>K</i>	11	1400
<i>Bng54</i>	<i>F</i>	11	1170	<i>Bng103</i>	<i>B</i>	3	1410
<i>Bng55</i>	<i>B</i>	5	1650	<i>Bng104</i>	<i>G</i>	1	1830
<i>Bng56</i>	<i>K</i>	6	2480	<i>Bng105</i>	<i>D</i>	15	1260
<i>Bng57</i>	<i>D</i>	21	2090	<i>Bng106</i>	<i>C</i>	17	2680
<i>Bng58</i>	<i>F</i>	4	2400	<i>Bng107a</i>	<i>A</i>	3	2560
<i>Bng60</i>	<i>A</i>	16	2290	<i>Bng107b</i>	<i>K</i>	5	2560
<i>Bng61</i>	<i>D</i>	20	1580	<i>Bng108</i>	<i>D</i>	8	1540
<i>Bng62a</i>	<i>K</i>	3	2310	<i>Bng109</i>	<i>K</i>	13	1930
<i>Bng62b</i>	<i>F</i>	11	2310	<i>Bng110</i>	<i>K</i>	7	1270
<i>Bng63</i>	<i>C</i>	6	1560	<i>Bng111</i>	<i>K</i>	1	1970
<i>Bng64</i>	<i>K</i>	7	1620	<i>Bng112</i>	<i>J</i>	10	820
<i>Bng65a</i>	<i>H</i>	6	2080	<i>Bng113</i>	<i>H</i>	4	660
<i>Bng65b</i>	<i>E</i>	7	2080	<i>Bng114</i>	<i>C</i>	14	1540
<i>Bng67</i>	<i>J</i>	10	2170	<i>Bng115</i>	<i>D</i>	10	1930
<i>Bng68</i>	<i>I</i>	3	2530	<i>Bng116</i>	<i>C</i>	16	1930
<i>Bng69</i>	<i>F</i>	11	1000	<i>Bng117</i>	<i>D</i>	16	1430
<i>Bng70</i>	<i>J</i>	1	1970	<i>Bng118</i>	<i>A</i>	6	1760
<i>Bng71</i>	<i>B</i>	7	2240	<i>Bng119</i>	<i>D</i>	18	2220
<i>Bng72</i>	<i>H</i>	12	3880	<i>Bng120</i>	<i>J</i>	10	1520
<i>Bng73</i>	<i>F</i>	1	2410	<i>Bng121</i>	<i>I</i>	6	1710
<i>Bng74</i>	<i>D</i>	19	650	<i>Bng122</i>	<i>H</i>	4	1330
<i>Bng75</i>	<i>C</i>	15	1600	<i>Bng123</i>	<i>C</i>	13	1180
<i>Bng76</i>	<i>J</i>	10	3120	<i>Bng124</i>	<i>C</i>	3	1270
<i>Bng77</i>	<i>D</i>	5	2500	<i>Bng125</i>	<i>F</i>	11	1140
<i>Bng78</i>	<i>J</i>	10	2270	<i>Bng126</i>	<i>H</i>	4	2510
<i>Bng79</i>	<i>K</i>	13	2610	<i>Bng128</i>	<i>F</i>	11	2340
<i>Bng80a</i>	<i>A</i>	1	1300	<i>Bng129</i>	<i>K</i>	13	1710
<i>Bng80b</i>	<i>G</i>	14	1300	<i>Bng130</i>	<i>H</i>	14	950
<i>Bng81</i>	<i>A</i>	3	1090	<i>Bng131</i>	<i>F</i>	2	1190
<i>Bng82</i>	<i>D</i>	10	1480	<i>Bng132</i>	<i>K</i>	13	1120
<i>Bng83</i>	<i>H</i>	1	870	<i>Bng133</i>	<i>E</i>	5	3120
<i>Bng84</i>	<i>D</i>	6	2190	<i>Bng134</i>	<i>K</i>	6	3830
<i>Bng86</i>	<i>C</i>	15	1650	<i>Bng135</i>	<i>A</i>	16	3000

Table 1. Continued.

Locus	Linkage group	Relative position	Insert size (bp)	Locus	Linkage group	Relative position	Insert size (bp)
<i>Bng136</i>	<i>J</i>	2	1190	<i>Bng183</i>	<i>G</i>	4	1210
<i>Bng137</i>	<i>G</i>	12	1710	<i>Bng184</i>	<i>B</i>	4	530
<i>Bng138</i>	<i>F</i>	9	1320	<i>Bng186</i>	<i>F</i>	11	1350
<i>Bng139</i>	<i>F</i>	12	2540	<i>Bng187</i>	<i>J</i>	4	1860
<i>Bng140</i>	<i>K</i>	6	1780	<i>Bng188</i>	<i>H</i>	8	1960
<i>Bng141</i>	<i>D</i>	4	1340	<i>Bng189</i>	<i>H</i>	4	2030
<i>Bng142a</i>	<i>C</i>	19	3470	<i>Bng190</i>	<i>D</i>	9	1330
<i>Bng142b</i>	<i>H</i>	11	3470	<i>Bng191</i>	<i>A</i>	22	2030
<i>Bng143</i>	<i>K</i>	8	950	<i>Bng192</i>	<i>J</i>	5	1300
<i>Bng144</i>	<i>D</i>	6	2220	<i>Bng193</i>	<i>G</i>	5	2620
<i>Bng145</i>	<i>J</i>	5	1930	<i>Bng195</i>	<i>H</i>	13	2700
<i>Bng146</i>	<i>A</i>	14	1550	<i>Bng197</i>	<i>K</i>	14	2560
<i>Bng148</i>	<i>D</i>	14	2220	<i>Bng198</i>	<i>K</i>	12	2140
<i>Bng149</i>	<i>K</i>	13	820	<i>Bng199</i>	<i>A</i>	19	1230
<i>Bng150</i>	<i>K</i>	2	1940	<i>Bng200</i>	<i>I</i>	1	2340
<i>Bng151</i>	<i>B</i>	8	1030	<i>Bng201</i>	<i>D</i>	4	1320
<i>Bng152</i>	<i>E</i>	6	1070	<i>Bng202</i>	<i>G</i>	1	1620
<i>Bng153</i>	<i>A</i>	15	1900	<i>Bng203</i>	<i>A</i>	10	2990
<i>Bng154</i>	<i>J</i>	10	1050	<i>Bng204</i>	<i>A</i>	12	1420
<i>Bng155</i>	<i>C</i>	8	2810	<i>Bng205a</i>	<i>F</i>	8	1350
<i>Bng156</i>	<i>K</i>	4	2500	<i>Bng205b</i>	<i>E</i>	7	1350
<i>Bng157</i>	<i>A</i>	7	1310	<i>Bng206</i>	<i>J</i>	10	1290
<i>Bng158</i>	<i>J</i>	10	2310	<i>Bng209</i>	<i>G</i>	10	1020
<i>Bng159</i>	<i>D</i>	3	3210	<i>Bng211a</i>	<i>A</i>	21	1810
<i>Bng160</i>	<i>B</i>	8	2280	<i>Bng211b</i>	<i>C</i>	1	1810
<i>Bng161</i>	<i>E</i>	9	1290	<i>Bng212</i>	<i>D</i>	16	1380
<i>Bng162</i>	<i>E</i>	2	590	<i>Bng213</i>	<i>F</i>	4	2180
<i>Bng163</i>	<i>K</i>	13	2180	<i>Bng214</i>	<i>F</i>	11	2330
<i>Bng164</i>	<i>C</i>	11	1480	<i>Bng215</i>	<i>G</i>	5	1350
<i>Bng165</i>	<i>C</i>	9	2260	<i>Bng216</i>	<i>C</i>	6	2040
<i>Bng166</i>	<i>E</i>	7	2170	<i>Bng218</i>	<i>I</i>	8	2020
<i>Bng167</i>	<i>K</i>	6	520	<i>Bng219</i>	<i>I</i>	4	580
<i>Bng168a</i>	<i>A</i>	22	2210	<i>Bng220</i>	<i>F</i>	11	4250
<i>Bng168b</i>	<i>K</i>	4	2210	<i>Bng221</i>	<i>C</i>	10	2100
<i>Bng170</i>	<i>A</i>	20	1170	<i>Bng222</i>	<i>A</i>	12	2480
<i>Bng171</i>	<i>H</i>	3	1900	<i>Bng223</i>	<i>A</i>	13	2350
<i>Bng172</i>	<i>I</i>	7	3120	<i>Bng224</i>	<i>B</i>	1	880
<i>Bng173</i>	<i>H</i>	2	2050	<i>Bng225</i>	<i>G</i>	11	1030
<i>Bng174</i>	<i>D</i>	12	2600	<i>Bng226</i>	<i>K</i>	13	2220
<i>Bng175</i>	<i>C</i>	13	880	<i>Bng227</i>	<i>H</i>	5	700
<i>Bng176</i>	<i>K</i>	12	3120	<i>Bng228</i>	<i>K</i>	9	1840
<i>Bng177</i>	<i>G</i>	7	930	<i>Bng230</i>	<i>K</i>	6	500
<i>Bng178</i>	<i>D</i>	13	1310	<i>Bng231</i>	<i>K</i>	13	2100
<i>Bng179</i>	<i>G</i>	2	2240	<i>Bng232</i>	<i>C</i>	7	1770
<i>Bng180</i>	<i>D</i>	17	2550	<i>Bng234</i>	<i>I</i>	2	900
<i>Bng181</i>	<i>K</i>	15	1770	<i>Bng235</i>	<i>E</i>	1	2030
<i>Bng182</i>	<i>J</i>	9	1180				

Table 2 List of protein marker loci isozymes and seed storage proteins

Isozyme locus	Linkage group	Relative position	References ^a	Seed protein locus	Linkage Group	Relative position	References
<i>Aco1</i>	<i>D</i>	–	1, 3	<i>AAI1</i>	<i>B</i>	2	3
<i>Aco2</i>	<i>E</i>	3	1, 3	<i>AAI2</i>	<i>B</i>	2	3
<i>Adh1</i>	<i>K</i>	9	1, 3	<i>Pha</i>	<i>A</i>	15	2, 3
<i>Bnag</i>	<i>F</i>	11	1, 3	<i>Spa</i>	<i>A</i>	9	2, 3
<i>Dial</i>	<i>E</i>	8	1, 3	<i>Spb</i>	<i>A</i>	9	2, 3
<i>Est2</i>	<i>A</i>	16	1, 3	<i>Spba</i>	<i>A</i>	5	2, 3
<i>Got2</i>	<i>K</i>	10	1, 3	<i>Spc</i>	<i>H</i>	6	2, 3
<i>Mdh1</i>	<i>A</i>	1	1, 3	<i>Spd</i>	<i>B</i>	2	2, 3
<i>Skdh</i>	<i>C</i>	18	1, 3	<i>Spe</i>	<i>A</i>	14	2, 3

^a1 Vallejos and Chase 1991a, 2 Vallejos and Chase 1991b, 3 Vallejos et al 1992

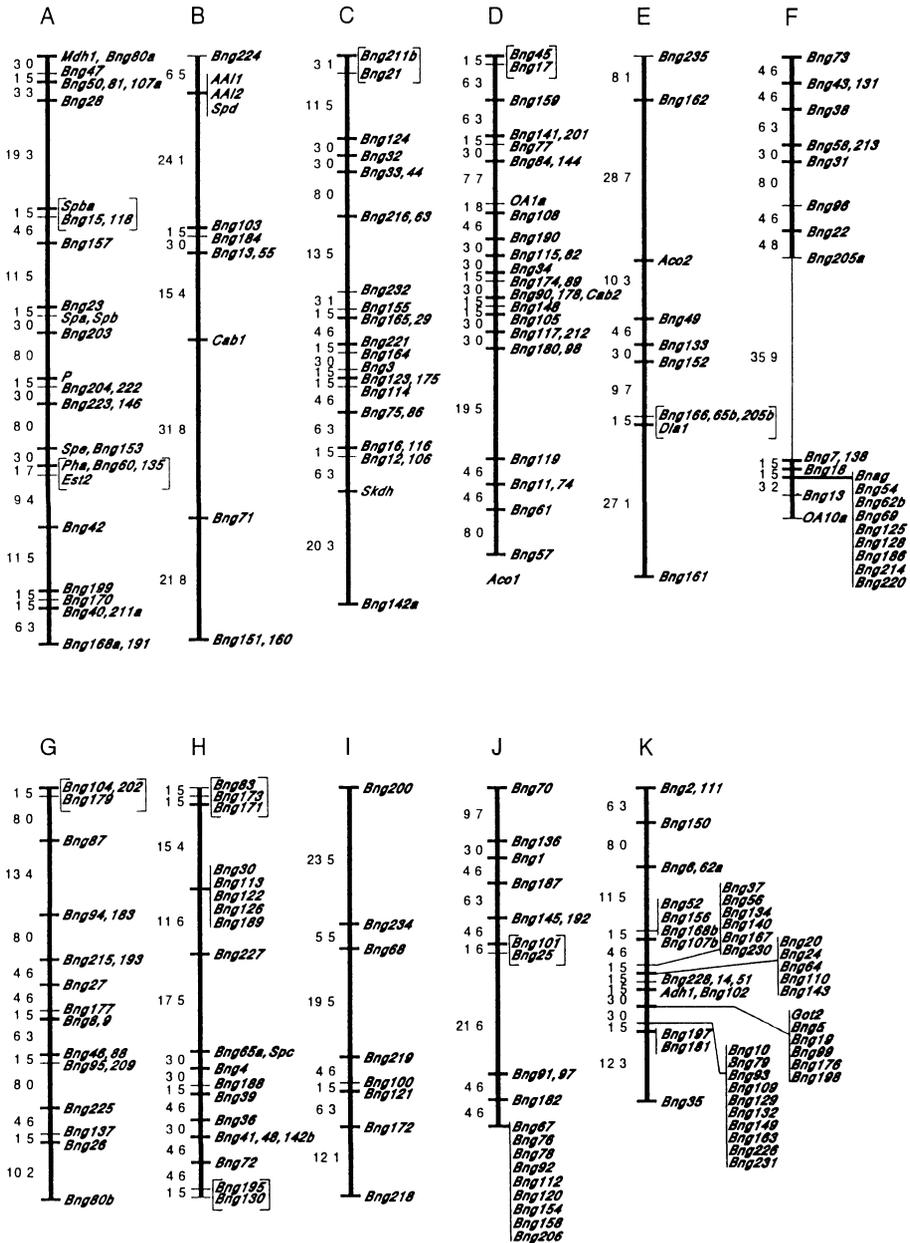
3. Some applications of RFLP markers in common bean

Although the current and potential uses of molecular markers have been treated in detail in the first chapter of this volume, I would like to point out three targets of interest in common beans: identification of genetic factors that affect gene flow between gene pools, analysis of QTLs that affect yield, and tagging genes for resistance to disease for plant breeding purposes and direct genomic cloning. Mesoamerica and the northern Andean region of South America have been identified as the two major centers of diversity (Gepts et al. 1986; Koenig and Gepts 1989). In addition to the group of landraces and modern cultivars that constitute the primary gene pool of common beans, additional genetic variation can be found in the secondary gene pool that comprises the wild forms of *P. vulgaris*, in the tertiary gene pool that corresponds to the *P. coccineus* complex, and finally in the quaternary gene pool that includes *P. acutifolius* and other species (Hidalgo 1991). These pools represent a valuable source of genes of economic importance. Unfortunately, gene transfer between the pools can be hampered by intra- and interspecific genetic barriers. For instance, differential photoperiodic responses within the andean gene pool alone can present some

→

Fig 1 Phaseolus vulgaris L linkage map *Mapmaker* (Lander et al 1987) was used to group and order 250 markers. These markers have been assigned to 150 loci/locus clusters on 11 linkage groups (A–K). The exact position of *Aco1* on group *D* could not be located due to incomplete data for this locus, *Aco1* is listed at the bottom of the group *A*. *framework* order was obtained after grouping the markers (LOD 4.0, 25 cM), and then ordering each group (LOD 3.0, 25 cM, three point exclusion LOD – 3.0). Markers that belong to the *framework* order have been marked with thicker crossbars. Loci for which an order could not be established (LOD >2.0) have been enclosed by brackets. Numbers on the left of a linkage group represent map distances in cM. Loci at distances that exceed 25 cM in groups *B*, and *E*, have an LOD score > 4.0 and have been included in the *framework*. The distance (35.9 cM) between *Bng205a* and *Bng7* (group *F*) was calculated using the Kosambi function. Although an LOD of 2.22 was found for this linkage, other independent data support their assignment to this group.

difficulties for within-pool gene transfer (Brücher 1988). Even more dramatic is the dwarf-lethal two gene system (DL₁ and DL₂) that restricts gene flow between some Mesoamerican and Andean accessions (Gepts and Bliss 1985; Shii et al. 1980; Singh and Gutierrez 1984). The existence of some genetic factors that affect compatibility in interspecific crosses has been suggested and it is supported by the identification of certain genotype combinations with increased compatibility (Hucl and Scoles 1985; Parker and Michaels 1986). The extent of



structural similarities in the chromosomes of different gene pool members can also affect the effectiveness of gene transfer. Only a modest characterization of the cytogenetics of the *Phaseolus* group has been achieved due to the small size of the chromosomes. It is not known for instance whether the chromosomes of closely related species are homosequential. For example, Cheng et al. (1981) reported two chromosome inversions on different chromosomes in an interspecific hybrid between *P. vulgaris* and *P. coccineus*; however these findings have been contested by Shii et al. (1982). Identification and molecular tagging of genetic factors that restrict gene flow between the gene pools of beans will permit the design of strategies to facilitate an effective gene transfer. The availability of the RFLP map in beans will also permit a comparison of the genomes of the secondary, tertiary and quaternary gene pools, a project that is currently underway in my laboratory. Comparisons of this kind have already been performed in the Solanaceae (Bonierbale et al. 1988; Tanksley et al. 1988).

Common beans lend themselves as a good model system to study QTLs that affect yield. Wallace and Masaya (1988) have developed a 'yield system analysis' to investigate the genetic components of yield and their interactions with the environment. In addition, Hoogenboon et al. (1988) have developed a computer simulation model for common beans - BEANGRO. The application of molecular markers to recombinant inbred lines (Burr et al. 1989), generated from suitable contrasting genotypes, will be useful in the identification of both genetic factors that affect yield and the responses these factors have to different environments. Molecular markers have been used to detect a number of QTLs in tomato (Paterson et al. 1988; see also Chapter 4 in this volume). This information can in turn be used to refine computer simulation programs that would take into account specific genetic factors. For instance, a negative correlation between seed size and yield has been reported for beans (Coyne 1968). A genetic factor that affects seed size has been identified via isozyme linkage analysis (Vallejos and Chase 1991a). Thus, tagging genes that affect the different components of yield can lead to the construction of specific genetic stocks carrying one or multiple combinations of these genes. These stocks can then be used to ask specific questions about the role certain genes play in different physiological processes such as sink-source relationships or photosynthate partitioning.

Finally, a large number of genes involved in disease resistance is available for molecular tagging in beans. There are at least 21 genetically characterized monogenic virus resistances (Provvidenti 1987), and a few other resistances to bacterial and fungal pathogens (Bassett 1989). Molecular tags can facilitate the efficient pyramiding of appropriate resistances into single breeding lines tailored for specific environments. Furthermore, the relatively small size of the bean genome opens the possibility of molecular cloning of any of these resistances via chromosome walking techniques (Rommens et al. 1989). The average ratio of physical distance to map distance has been estimated at 530 kb/cM in beans (Vallejos et al. 1992). Development of high density maps around a specific disease resistance gene will expedite the isolation of its genomic clone.

4. References

- Arumuganatham, K and Earle, E D (1991) Nuclear DNA content of some important plant species *Plant Mol Biol Rep* 9 208–218
- Bassett, M J (1989) List of genes – *Phaseolus vulgaris* L *Annu Rept Bean Improv Coop* 32 1–7
- Bassett, M J (1991) A revised linkage map of common bean *HortScience* 26 834–836
- Bonierbale, M, Plaisted, R and Tanksley, S D (1988) Construction of comparative genetic maps of potato and tomato based on a common set of cloned sequences *Genetics* 120 1095–1103
- Brucher, H (1988) The wild ancestor of *Phaseolus vulgaris* in South America In P Gepts (ed), *Genetic Resources of Phaseolus Beans*, pp 185–214 Kluwer Academic Publishers, Dordrecht
- Burr, B, Burr, F A, Thompson, K H, Albertson, M C and Stuber, C W (1988) Gene mapping with recombinant inbreds in maize *Genetics* 118 519–526
- Cheng, S S, Bassett, M J and Quesenberry, K H (1981) Cytogenetic analysis of interspecific hybrids between common bean and scarlet runner bean *Crop Sci* 21 75–79
- Chase, C D, Ortega, V M and Vallejos, C E (1991) DNA restriction fragment length polymorphisms correlate with isozyme diversity in *Phaseolus vulgaris* L *Theor Appl Genet* 81 806–811
- Coyne, D P (1968) Correlation, heritability, and selection of yield components in field beans, *Phaseolus vulgaris* L *Proc Am Soc Hort Sci* 93 388–396
- Gepts, P and Bliss, F A (1985) F₁ hybrid weakness in the common bean: differential geographic origin suggests two gene pools in cultivated germplasm *J Hered* 76 447–450
- Gepts, P, Osborn, T C, Rashka, K and Bliss, F A (1986) Phaseolin protein variability in wild forms and land races of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication *Econ Bot* 40 451–468
- Hidalgo, R (1991) CIAT's world *Phaseolus* collection In A van Schoonhoven and O Voysest (eds), *Common Beans: Research for Crop Improvement*, pp 163–197 C A B International, Wallingford, U K
- Hoogenboom, G, Jones, J W, White, J W and Boote, K J (1988) BEANGRO V 1 0 Dry Bean Crop Growth Simulation Model User's Guide Department of Agricultural Engineering, University of Florida
- Hucl, P and Scoles, G J (1985) Interspecific hybridizations in the common bean: a review *HortScience* 20 352–357
- Koeng, R and Gepts, P (1989) Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of genetic diversity *Theor Appl Genet* 78 809–817
- Lander, E S, Green, P, Abrahamson, J, Baarlow, A, Daly, M J, Lincoln, S E and Newburg, L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations *Genomics* 1 174–181
- Mendel, G (1866) *Experiments in Plant Hybridisation* Translated by the Royal Horticultural Society of London in 1938 Harvard University Press, Cambridge, MA
- Moreno, J and Chrispeels, M J (1989) A lectin gene encodes the α -amylase inhibitor of the common bean *Proc Natl Acad Sci U S A* 86 7885–7889
- Parker, J P and Michaels, T E (1986) Simple genetic control of hybrid plant development in interspecific crosses between *Phaseolus vulgaris* L and *P acutifolius* A Gray *Plant Breed* 97 315–323
- Paterson, A H, Lander, E S, Hewit, J D, Peterson, T S, Lincoln, S E and Tanksley, S D (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms *Nature* 335 721–726
- Provvidenti, R (1987) List of genes in *Phaseolus vulgaris* for resistance to viruses *Annu Rept Bean Improv Coop* 30 1–4
- Rommens, J M, Iannuzzi, M C, Kerem, B-S, Drumm, M L, Melmer, G, Dean, M, Rozmahel, R, Cole, J L, Kennedy, D, Hidaka, N, Zsiga, M, Buchwald, M, Riordan, J R, Tsui, L-C and Collins, F S (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping *Science* 245 1059–265

- Shaw, J K and Norton, J B (1918) The inheritance of seed-coat color in garden beans Massachusetts Agric Exp Sta Bull 185 58-104
- Shu, C T, Rabakoarihanta, A, Mok, M C and Mok, D W S (1982) Embryo development in reciprocal crosses of *Phaseolus vulgaris* L and *P. coccineus* Lam Theor Appl Genet 62 59-64
- Sun, S M, Slightom, J L and Hall, T C (1981) Intervening sequence in a plant gene - comparison of the partial sequence of cDNA and genomic DNA of French bean phaseolin Nature 289 37-41
- Talbot, D R, Adang, M J, Slighton, J L and Hall, T C (1984) Size and organization of a multigene family encoding phaseolin, the major seed storage protein of *Phaseolus vulgaris* L Mol Gen Genet 198 42-49
- Tanksley, S D, Bernatzky, R, Lapitan, L L and Prince, J P (1988) Conservation of gene repertoire but not gene order in pepper and tomato Proc Natl Acad Sci U S A 85 6419-6423
- Thompson, W F, Everett, M, Polans, N O and Jorgensen, R A (1983) Phytochrome control of RNA levels in developing pea and mungbean leaves Planta 158 487-500
- Vallejos, C E and Chase, C D (1991a) Linkage between isozyme markers and a locus affecting seed size in *Phaseolus vulgaris* L Theor Appl Genet 81 413-419
- Vallejos, C E and Chase, C D (1991b) Extended linkage map for the phaseolin linkage group of *Phaseolus vulgaris* L Theor Appl Genet 82 353-357
- Vallejos, C E, Sakiyama, N S and Chase, C D (1992) A molecular marker-based linkage map of *Phaseolus vulgaris* L Genetics 131 733-740
- Wallace, D H and Masaya, P N (1988) Using yield trial data to analyze the physiological genetics of yield accumulation and the genotype X environment interaction effects on yield Annu Rep Bean Improv Coop 31 vii-xxiv
- Williams, J G K, Kubelik, A R, Livak, K J, Rafalski, J A and Tingey, S V 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers Nucl Acids Res 18 6531-6535
- Zheng, J, Nakata, M, Uchiyama, H, Morikawa, H and Tanaka, R (1991) Giemsa C-banding pattern in several species of *Phaseolus* L and *Vigna* Savi, fabacea Cytologia 56 459-466

16. RFLP map of the potato

CHRISTIANE GEBHARDT, ENRIQUE RITTER¹ and FRANCESCO SALAMINI

Max-Planck-Institut für Züchtungsforschung, Carl-von-Linné-Weg, D-50829 Köln, Germany,

¹Granja Modelo - CIMA, km 356 de la N I, Arkaute (Alava), Apartado Correos 46 E-01080 Vitoria-Gasteiz, Spain

Contents

1	Introduction	271	5 2	GP-markers	274
2	The mapping population	272	5 3	Marker loci identified by functional sequences	274
3	The experimental system	272	5 4	TG-markers	274
4	Map construction	273	5 5	Morphological markers	274
5	Nomenclature and origin of RFLP markers	273	6	The RFLP map of potato	276
5 1	CP-markers	273	7	References	283

1. Introduction

The potato, *Solanum tuberosum*, a species of the family Solanaceae, is cultivated in most temperate and subtropical zones of the world. After wheat, rice and corn it occupies the fourth position in terms of world production per year (FAO Yearbook 1988).

The history of the potato and its impact during the development of the industrial society has been described in detail by Salaman (1985). Most likely, the potato originated in the Andean region of South-America where it has been cultivated for at least 2000 years. After the Spanish conquest, the potato was introduced into Europe via Spain and England in the 16th century. Historical evidence indicates that the first potato tubers reaching the European continent were botanically *Solanum tuberosum* subsp. *andigena*, producing tubers only under short day photoperiodic conditions. The conversion from an exotic ornamental plant being of interest only to few botanists into a major food crop supporting the growing population in the industrial society of the 18th and 19th centuries, was made possible by natural or unintentional selection of genotypes producing tubers under the long day photoperiod of Europe. The European potato of today has the botanical name *Solanum tuberosum* subsp. *tuberosum*. It is a tetraploid plant with 48 chromosomes and tetrasomic inheritance. One genome complement has, therefore, twelve chromosomes.

As potatoes are vegetatively propagated via tubers, there was no need to select for highly fertile genotypes during potato breeding. This fact, together with the tetrasomic inheritance of the crop, prevented the development of genetic linkage maps. Only a few linkages were reported, for example among anthocyanin pigmentation genes and between pigmentation and tuber shape (reviewed by De Jong 1991). Centromere map distances of isozyme loci were determined by Douches and Quiros (1987). Reduction of the ploidy from the tetraploid to the diploid level is possible either by pollination of tetraploid genotypes with certain diploid strains of *Solanum phureja* which induced the parthenogenetic development of diploid gametes into plants (Hougas et al. 1964; Hermsen and Verdenius 1973), or by regenerating plants from diploid male gametes via anther or microspore culture of tetraploid parents (Dunwell and Sunderland 1973; Powell and Uhrig 1987). Diploid potatoes are, however, largely self-incompatible. This fact, and the high genetic load present in the species make the construction of pure lines in most cases impractical. The RFLP map of potato is based, therefore, on segregating progeny of highly heterozygous diploid parents. The linkage map is, of course, useful for understanding the genome of tetraploid potatoes.

2. The mapping population

Two diploid, heterozygous *S. tuberosum* subsp. *tuberosum* breeding lines were crossed to give a F_1 . The heterozygosity of the parents was estimated to be 57% and 59%, respectively, based on RFLP alleles. The parents exhibited 82% informative polymorphisms among each other when compared with 147 DNA probes and three restriction enzymes tested per probe. Under the given experimental conditions (Gebhardt et al. 1989) this value was close to the mean value (80%) found in a gene pool of 38 diploid potato genotypes (Gebhardt et al. 1989). Pollinating an individual F_1 plant with one of the parents yielded a backcross progeny of 67 lines (Gebhardt et al. 1989, 1991). Parents and progeny were clonally propagated via tubers and are available from the BGRC *in vitro* collection (Bundesforschungsanstalt für Landwirtschaft, Braunschweig-Völkenrode, FAL, Bundesallee 50, D-3300 Braunschweig, Germany).

3. The experimental system

Details of the experimental system by which the segregation data for the potato RFLP map were obtained are described by Gebhardt et al. (1989). In short, total genomic DNA was isolated from freeze-dried leaves and shoots of plants grown in the greenhouse under normal daylight conditions. The DNA was digested with the four base cutter restriction enzymes *TaqI*, *RsaI* and *AluI*, respectively. Approximately 55% of the segregation data were obtained with *TaqI*-, 35% with *RsaI*- and only 10% with *AluI*-restricted DNA. Restriction fragments were

separated on 4% polyacrylamide gels under denaturing conditions and transferred to nylon membranes by electroblotting. The separation range was between 250 and 2000 bases resolving minimum length differences of ca. 5 bases. The membranes were hybridized to ^{32}P labelled probes and washed at a moderate stringency.

4. Map construction

The principles and algorithms on which the RFLP segregation data were converted into linkage groups have been described by Ritter et al. (1990). Whereas a single genetic model applies to segregating alleles of F_2 - or backcross-mapping populations derived from pure lines, in populations derived from partially heterozygous parents RFLP alleles segregate according to several different genetic models. In our backcross type mapping population, between one and three RFLP alleles segregated per locus. Autoradiographs were evaluated by scoring presence versus absence of segregating individual restriction fragments without taking into consideration possible allelism among different fragments. No distinction was made between a fragment A in the homozygous (AA) or heterozygous (AO) state. In scoring RFLPs phenotypically, three types of segregation patterns were distinguished: fragments being present in the heterozygous state either in the female or in the male parent segregating with a 1:1 ratio (presence versus absence, testcross type segregation); fragments being present in the heterozygous state in both parents segregating with a 3:1 ratio (presence versus absence, F_2 type segregation). Recombination frequencies among all fragments scored over all markers tested were estimated using algorithms proper for the genetic models fitting the observed segregations (Ritter et al. 1990). Linkage subgroups were constructed based on the recombination frequencies found among fragments segregating with the same ratio and, in the case of the 1:1 segregations, separately for the female and male parent. The linkage subgroups were connected and oriented relative to each other considering RFLP loci with allelic fragments belonging to at least two linkage subgroups. Allelism among RFLP fragments detected by the same probe was assumed when they were found linked with zero percent recombination (in coupling or repulsion phase). Following this strategy, three linkage groups were obtained for each chromosome: one for the female and one for the male parent (testcross), and the third resulting from both parents (F_2).

5. Nomenclature and origin of RFLP markers

5.1. CP-markers

The RFLP map contains 84 marker loci based on random cDNA sequences of potato cloned in the Bluescript vector (Stratagene). The source tissues for the

cDNA library were young leaves and shoots harvested from several diploid potato genotypes. Selection criteria were – besides polymorphism – the insert size (> 200 bp) and pattern complexity (low repetitiveness) (Gebhardt et al. 1989). Each cDNA marker locus is identified by the letters CP followed by an identification number. Multiple loci detected with the same marker probe are indicated by small letters a, b, c, ... in parentheses behind the identification number. Of 65 cDNA probes mapped, 52 (80%) detected a single locus, 11 (17%) detected sequences at two loci, one at 4 and one at 6 loci.

5.2. GP-markers

The majority of marker loci (246) are based on 186 random genomic sequences of potato. *PstI* digested total genomic DNA was cloned in the Bluescript vector and selected for size (500–2000 bp) and low copy number (Gebhardt et al. 1989). This marker class is indicated by the letters GP. Otherwise the nomenclature is the same as for CP markers. 148 (79%) genomic probes detected a single locus in the mapping population used, 29 (16%) two loci and 9 (5%) between 3 and 9 loci.

5.3. Marker loci identified by functional sequences

The map positions of 22 genes of known or operationally defined function are included in the RFLP map. The genes, their locus names, map positions, sources and references are listed in Table 1. All markers originated from potato with the exception of ShM which was derived from maize. Four genes, *Sr1*, *AmyZ1*, *TPT* and *PHA1* are included which were mapped using a population resulting from a different cross (Gebhardt et al. 1991), because they did not segregate in the progeny on which the map shown in Fig. 1 is based. The loci *PAT(a)* and *chitinase (a)* are clusters of genes encoding the patatin and chitinase multigene families (Ganal et al. 1991, L. Beerhus and E. Kombrink, pers. comm.).

5.4. TG-markers

Twenty six reference RFLP markers of the twelve homoeologous tomato chromosomes (Bonierbale et al. 1988) were included in the RFLP map of potato, therewith aligning the genetic maps of potato and tomato (Gebhardt et al. 1991). The markers are indicated by TG followed by an identification number. The marker probes were obtained from S.D. Tanksley (Cornell University, Ithaca, New York, U.S.A.).

5.5. Morphological markers

One morphological trait segregated in the mapping population with a 1:1 ratio. This was the tuber skin color (purple versus colorless). The locus is situated on chromosome X and was named *PSC* (purple skin color).

Table 1 Non-random DNA-sequences mapped

Specification	Marker name	No of loci	Map position	Sources and references
4-Coumarate CoA ligase	<i>4CL</i>	3	III(a), VI(b), XII(c)	(3), (1), (2)
Phenylalanine ammonia-lyase	<i>PAL</i>	5	IX(a,b,c), X(d e)	(3) (1), (2)
1,3 β -Glucanase	<i>Glucanase</i>	1	I	(4), (2)
Chitinase	<i>chitinase</i>	2	X(a,b)	(4) this chapter
Induced after infection of potato leaves with <i>Phytophthora infestans</i>	<i>PC116</i>	1	IV	(5) (1), (2)
PR1 (pathogenesis-related)	<i>prp1</i> (<i>pl471</i>)	1	IX	(5), (1), (2)
Ribulose biphosphate carboxylase small subunit	<i>rbcS</i>	3	II(<i>rbcS-c rbcS-2</i>), III(<i>rbcS-1</i>)	(6), (7) (1), (2)
Actin	<i>Actin</i>	1	V	(8), (1), (2)
Patatin	<i>PAT</i>	2	VIII(a), IV(b)	(9), (1), (2)
Wun-1 (wound-induced)	<i>WUN1</i>	1	VI	(10), (2)
Wun-2 (wound-induced)	<i>WUN2</i>	1	VI	(10), (2)
S-locus	<i>Sr1</i>	1	I	(11), (2)
Granule bound starch synthase	<i>WX</i>	1	VIII	(12), (1), (2)
Sucrose synthase	<i>ShM</i>	1	XII	(13), (2)
α -Amylase	<i>AmyZ3/4</i>	1	IV	(14), this chapter
	<i>AmyZ1</i>	1	IV	.. , .
Branching enzyme	<i>PBE</i>	1	IV	(15), this chapter
ADP-glucose pyrophosphorylase B	<i>AGPase B</i>	2	VII(a), XII(b)	(16), this chapter
ADP-glucose pyrophosphorylase S	<i>AGPase S</i>	2	I(a), VIII(b)	(16), this chapter
Plasma membrane H ⁺ -ATPase	<i>PHA1</i>	2	III(a), VI(b)	(17), this chapter
	<i>PHA2</i>	1	VII	, ..
Triosephosphate 3-phosphoglycerate translocator	<i>TPT</i>	1	X	(18), this chapter

(1) Gebhardt et al (1989), (2) Gebhardt et al (1991), (3) Fritzsche et al (1987), (4) L Beerhus and E Kombrink (pers comm , MPI fur Zuchtungsforchung Koln), (5) Taylor et al (1990), (6) Eckes et al (1985), (7) Wolter et al (1988), (8) M Thangavelu (pers comm , Plant Breeding Institute, Cambridge, U K), (9) Rosahl et al (1986), (10) Logemann et al (1988), (11) Kaufmann et al (1991), (12) Hergersberg (1988), (13) Werr et al (1985), (14) J Kreiberg (pers comm , Maribo Seed, Copenhagen, Denmark), (15) Koßmann et al (1991), (16) Muller-Röber et al (1990), (17) R V Wohner and W B Frommer (pers comm , Inst fur Genbiol Forschung, Berlin), (18) B Schulz and W B Frommer (pers comm , Inst fur Genbiol Forschung, Berlin)

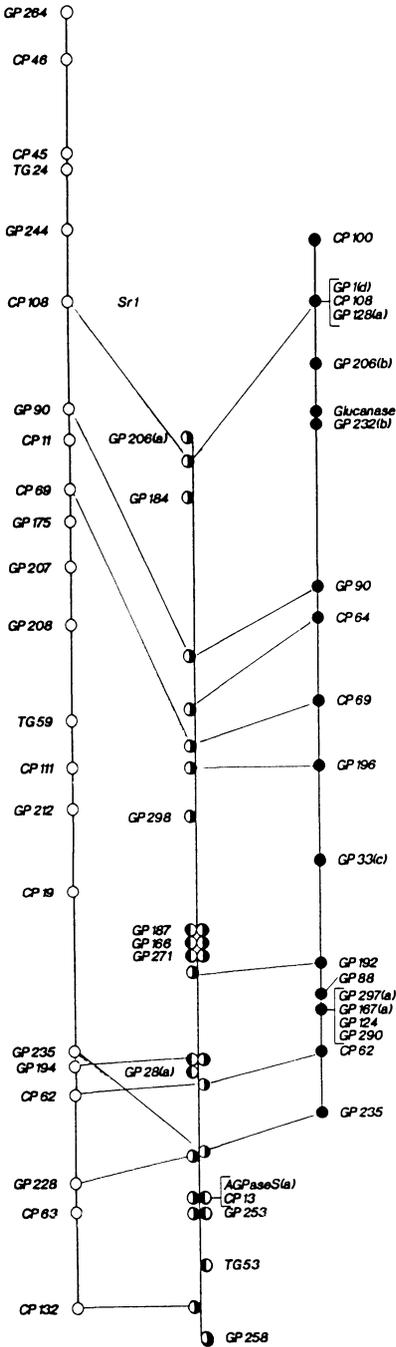
6. The RFLP map of potato

The genetic map of potato is shown in Fig. 1. Each chromosome is split into three linkage groups, two for the male and female parent, respectively, being equivalent to backcross type segregation analysis. The third linkage group is based on F_2 type segregation data. The chromosome nomenclature is the same as in the homoeologous tomato genome (see Chapter 20). 299 DNA markers and one morphological marker (tuber skin color) identified 384 loci. The percentage of duplicated RFLP loci is likely to be higher as not all loci present segregate in a single population. Twenty one percent of the markers detect two or more loci. Only few duplicated linkage blocks involving at least three markers have been found so far: *rbcS-c*, *CP70(b)* and *GP176(a)* on chromosome II are duplicated on the same chromosome (*GP176(b)*, *rbcS-2*, *CP70(a)*). Interestingly, a chromosome segment marked by *GP1(c)*, *CP15(b)*, *rbcS-c* on chromosome II seems to be duplicated on chromosome III (*rbcS-1*, *CP15(a)*, *GP1(a)*). The three *rbcS* loci present in the potato genome might therefore be the result of an ancient duplication of a chromosomal progenitor segment. A second duplicated linkage block involving four markers is detectable on chromosomes IX and XII (*GP91(a)*, *GP263(b)*, *CP20(a)*, *GP167(b)* on IX and *CP20(b)*, *GP263(a)*, *GP91(c)*, *GP167(c)* on XII). Unlike in maize (Helentjaris et al. 1988) and in *Brassica* (Slocum et al. 1990), it appears that there was no extensive duplication of chromosomes or chromosomal segments during the evolution of the potato genome.

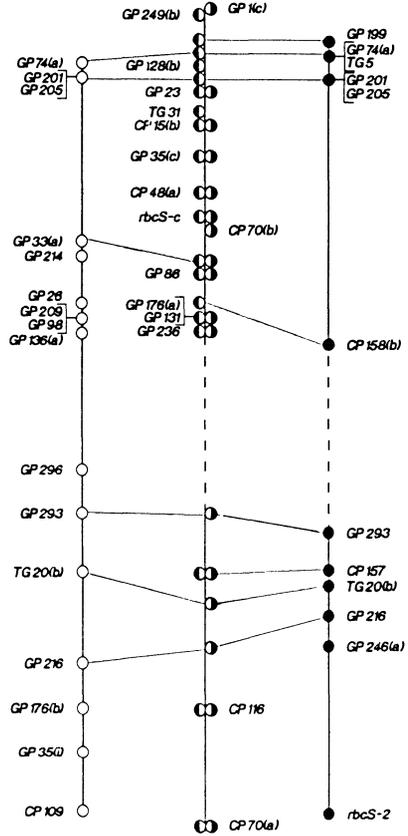
The length of the potato map with 384 loci, as derived from the intraspecific mapping population used, stands at ca. 1050 cM (units defined as by Kosambi 1944). This is not a significant increase as compared to the 1034 cM of a previously published map (Gebhardt et al. 1991) with 304 loci. Genome

→
 Fig 1 RFLP map of the potato. The twelve chromosomes are split into three linkage groups. The linkage groups on the left were obtained with restriction fragments descending from the male parent and segregating with a 1:1 ratio (open circles). The linkage groups on the right were deduced from restriction fragments descending from the female parent and also segregating with a 1:1 ratio (solid circles). The central linkage groups were obtained with fragments descending from both parents and segregating with a 3:1 ratio (half open, half solid circles). Pairs of circles in the central groups indicate allelism between two common fragments. Restriction fragments in the central groups which are linked to each other in coupling are shown as half circles in the same orientation. The half circles are shown inverted relative to each other when the fragments were in repulsion or coupling/repulsion phase (see Ritter et al. 1990). The lines connecting the linkage groups indicate that those restriction fragments belong to alleles of the same RFLP locus (allellic bridges). The orientation of the linkage subgroups relative to each other is determined by these allelic bridges. Marker fragments without significant linkage detectable among them but nevertheless on the same chromosome are connected by a broken line. Map distances are given in Centimorgan (Kosambi 1944). Marker loci resulting from genomic clones are indicated by GP, from cDNA clones by CP numbers. Marker loci derived from characterized sequences are shown with their specific names (Table 1). The map position of *Sr1* (I), *AmyZ1* (IV), *PHAI(a)* (III), *PHAI(b)* (VI) and *TPT* (X) was determined in a different cross. Small letters in parentheses indicate multiple loci detected by the same probe.

I



II



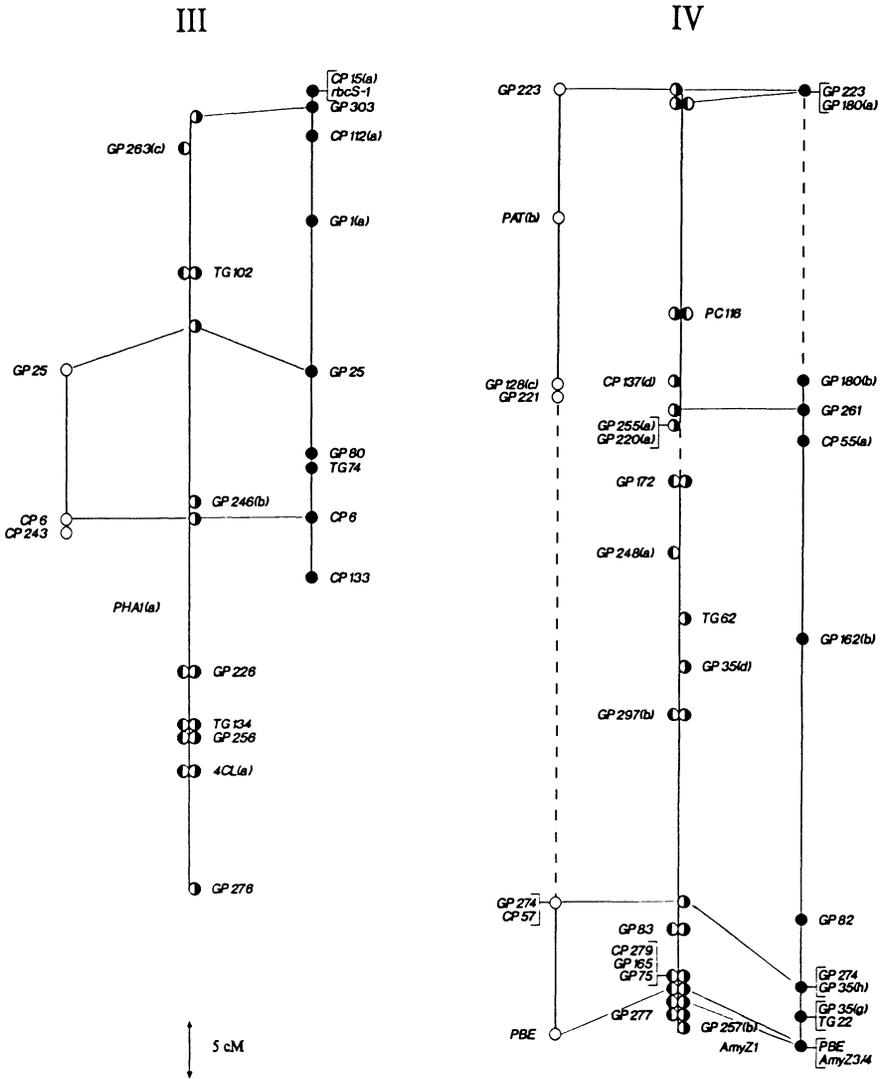


Fig. 1. Continued.

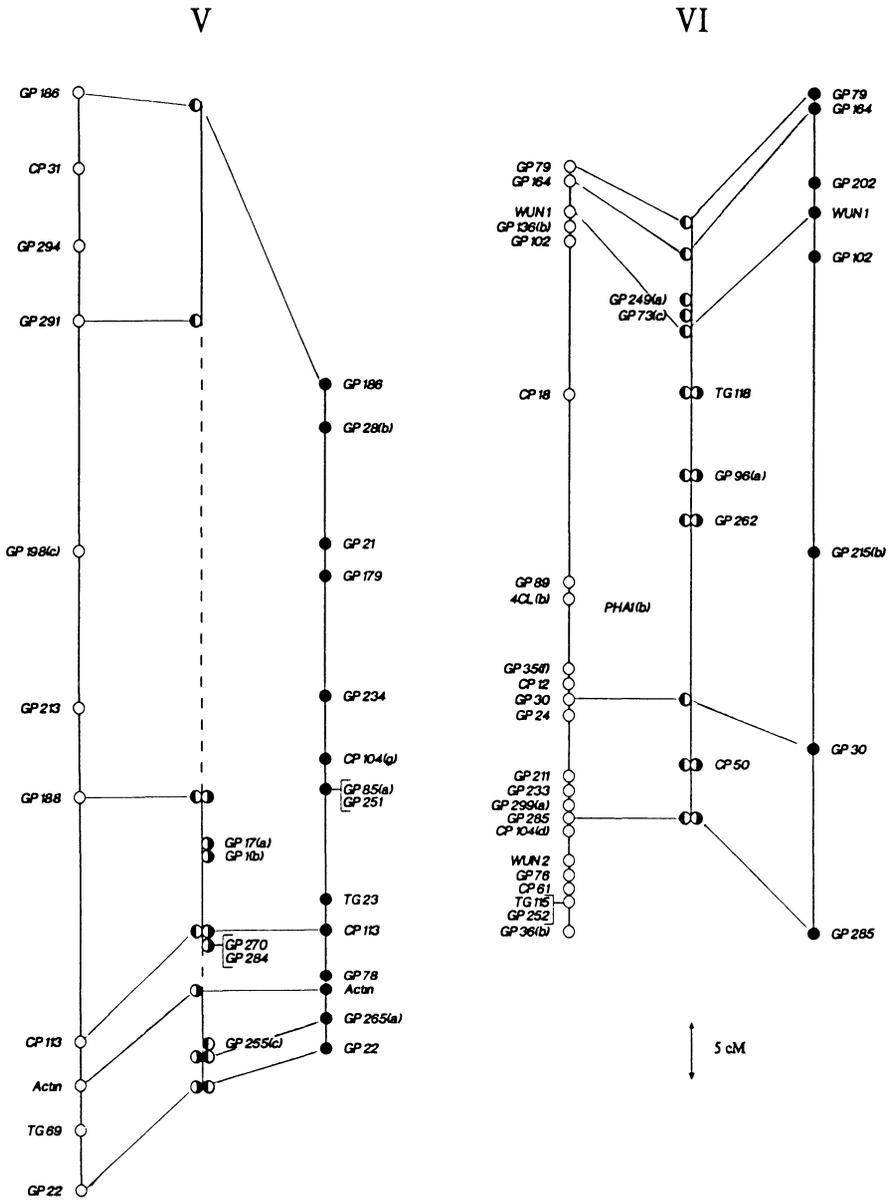


Fig 1 Continued

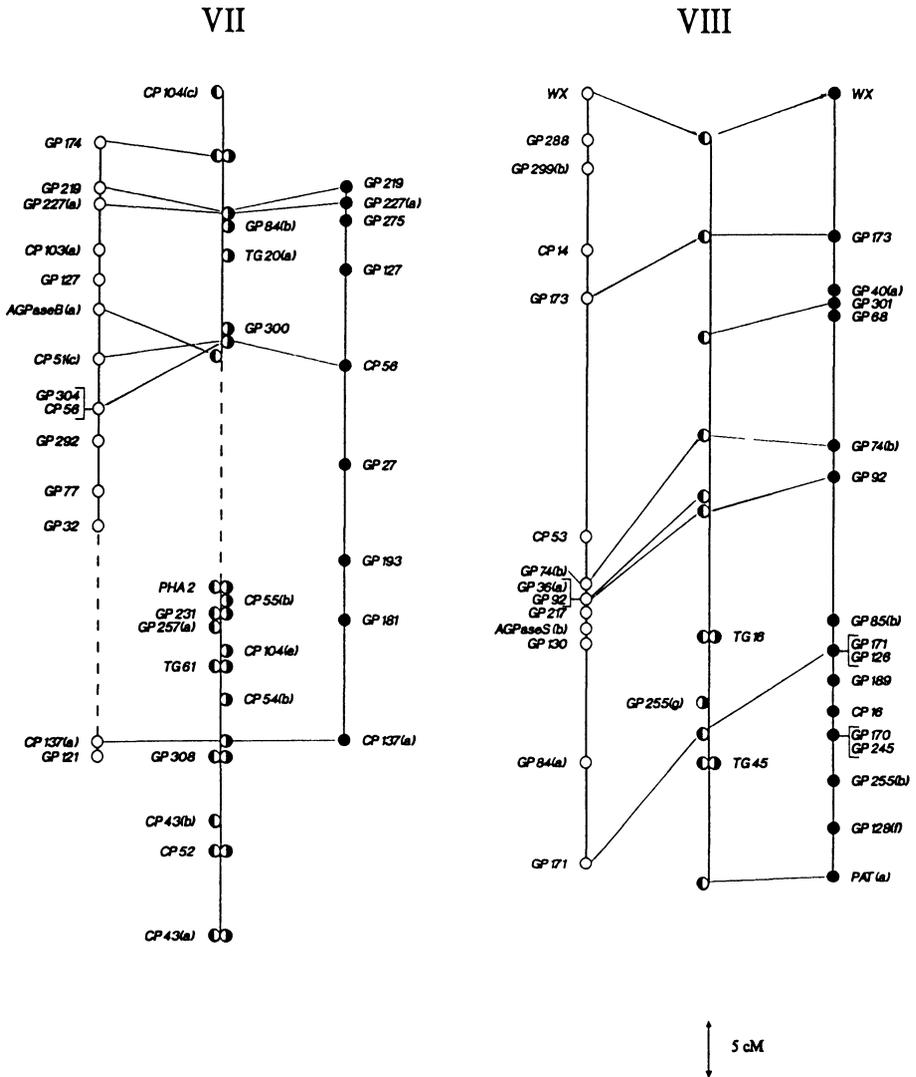


Fig. 1. Continued.

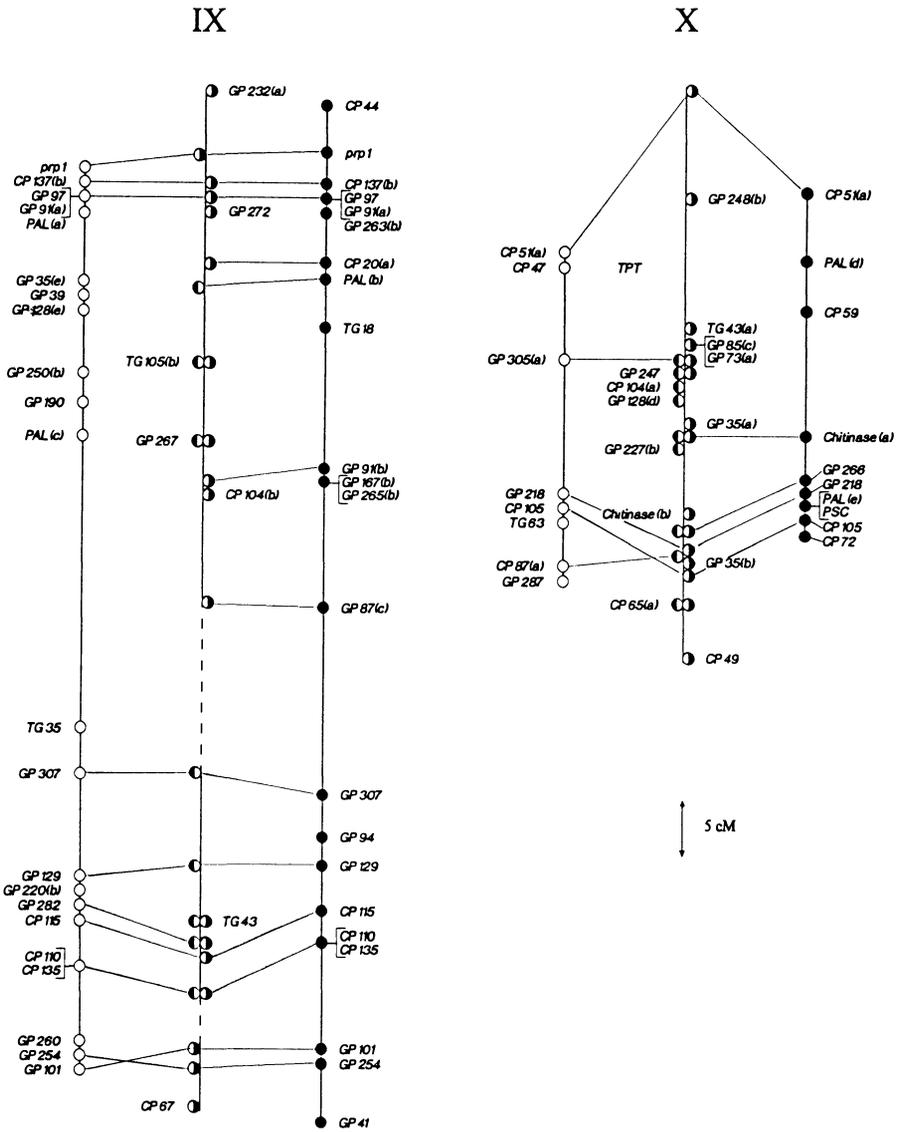


Fig. 1. Continued.

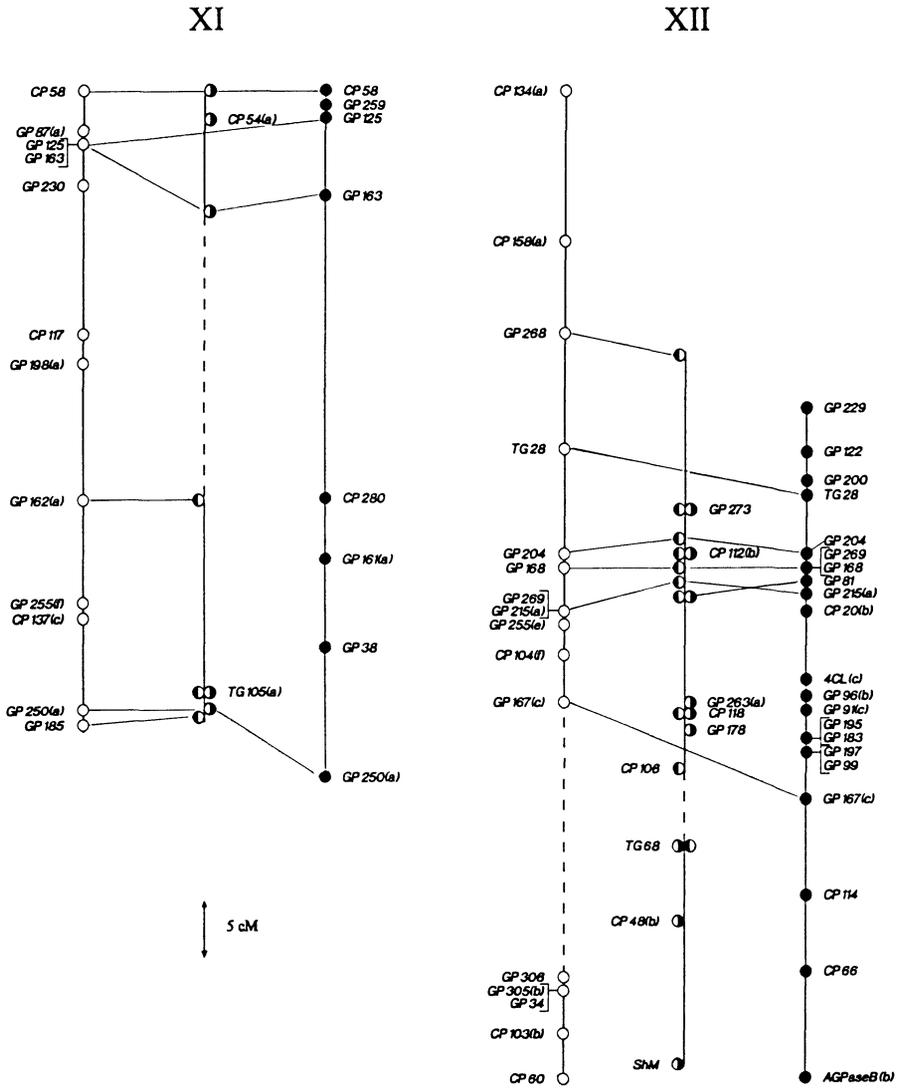


Fig. 1. Continued.

coverage by the map is estimated to be at least 90%. This is also supported by the observation that only three marker probes from more than 300 tested did not show significant linkage to any of the twelve established linkage groups.

Linkage maps derived from different genetic backgrounds in potato vary in map length (Bonierbale et al. 1989; Gebhardt et al. 1991). The order of loci, however, is conserved within the limits of the standard error associated with each map position (Gebhardt et al. 1991).

Segregation ratios deviating significantly from the expected 1:1 and 3:1 ratios, respectively, were found for 27% of the segregating restriction fragments. Several chromosomal segments with distorted segregation ratios were identified by linked distorted marker loci. Some of these segments were detectable in different genetic backgrounds, while others were specific for certain parents in particular cross combinations. The strongest distortion of segregation ratios was found on chromosome I and is assumed to be caused by the activity of the self-incompatibility locus (Gebhardt et al. 1991).

The RFLP map of potato is informative for other intra- and interspecific crosses. Mapped probes have been used in different genetic backgrounds to locate dominant resistance loci acting against the root cyst nematode *Globodera rostochiensis* (Barone et al. 1990; Gebhardt et al. 1993), potato virus X (Ritter et al. 1991), and the fungus *Phytophthora infestans* (Leonards-Schippers et al. 1992). The markers are also suitable for mapping quantitatively inherited traits in potato (unpublished results from this laboratory).

7. References

- Barone, A., Ritter, E., Schachtschabel, U., Debener, T., Salamini, F. and Gebhardt, C. (1990) Localization by restriction fragment length polymorphism mapping in potato of a major dominant gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. *Mol Gen Genet* 224 177–182
- Bonierbale, M W., Plaisted, R L. and Tanksley, S D. (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* 120 1095–1103
- De Jong, H. (1991) Inheritance of anthocyanin pigmentation in the cultivated potato. A critical review. *Am Potato J* 68 585–593
- Douches, D S. and Quiros, C F. (1987) Use of 4x-2x crosses to determine gene-centromere map distance of isozyme loci in *Solanum* species. *Genome* 29 519–527
- Dunwell, J M. and Sunderland, N. (1973) Anther culture of *Solanum tuberosum* L. *Euphytica* 22 317–323
- Eckes, P., Schell, J. and Willmitzer, L. (1985) Organ-specific expression of three leaf/stem specific cDNAs from potato is regulated by light and correlated with chloroplast development. *Mol Gen Genet* 199 216–224
- FAO Yearbook (1988) Production Vol 42, FAO Statistics Series No 88, Rome
- Fritzscheier, K H., Cretin, C., Kombrink, E., Rohwer, F., Taylor, J., Scheel, D. and Hahlbrock, K. (1987) Transient induction of phenylalanine ammonia-lyase and 4-coumarate CoA ligase mRNAs in potato leaves infected with virulent or avirulent races of *Phytophthora infestans*. *Plant Physiol* 85 34–41
- Ganal, M W., Bonierbale, M W., Roeder, M S., Park, W D. and Tanksley, S D. (1991) Genetic and physical mapping of the patatin genes in potato and tomato. *Mol Gen Genet* 225 501–509

- Gebhardt, C , Ritter, E , Debener, T , Schachtschabel, U , Walkemeier, B , Uhrig, H and Salamini, F (1989) RFLP analysis and linkage mapping in *Solanum tuberosum* Theor Appl Genet 78 65-75
- Gebhardt, C , Ritter, E , Barone, A , Debener, T , Walkemeier, B , Schachtschabel, U , Kaufmann, H , Thompson, R D , Bonierbale, M W , Ganai, M W , Tanksley, S D and Salamini, F (1991) RFLP maps of potato and their alignment with the homoeologous tomato genome Theor Appl Genet 83 49-57
- Gebhardt, C , Mugniery, D , Ritter, E , Salamini, F and Bonnel, E (1993) Identification of RFLP markers closely linked to the H1 gene conferring resistance to *Globodera rostochiensis* in potato Theor Appl Genet 85 541-544
- Helentjaris, T , Weber, D and Wright, S (1988) Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms Genetics 118 353-363
- Hergersberg, M (1988) Molekulare Analyse des waxy Gens aus *Solanum tuberosum* und Expression von antisense RNA in transgenen Kartoffeln Ph D thesis, University of Cologne
- Hermesen, J G Th and Verdenius, J (1973) Selection from *Solanum tuberosum* group phureja of genotypes combining high-frequency haploid induction with homozygosity for embryo spot Euphytica 22 244-259
- Hougas, R W , Peloquin, S J and Gabert, A C (1964) Effect of seed parent and pollinator on the frequency of haploids in *Solanum tuberosum* Crop Sci 4 593-595
- Kaufmann, H , Salamini, F and Thompson, R D (1991) Sequence variability and gene structure at the self-incompatibility locus of *Solanum tuberosum* Mol Gen Genet 226 457-466
- Kosambi, D D (1944) The estimation of map distances from recombination values Ann Eugen 12 172-175
- Koßmann, J , Visser, R G F , Muller-Rober, B T , Willmitzer, L and Sonnewald, U (1991) Cloning and expression analysis of a potato cDNA that encodes branching enzyme evidence for co-expression of starch biosynthetic genes Mol Gen Genet 230 39-44
- Leonards-Schippers, C , Gieffers, W , Salamini, F and Gebhardt, C (1992) The R1 gene conferring race-specific resistance to *Phytophthora infestans* in potato is located on potato chromosome I Mol Gen Genet 233 278-283
- Logemann, J , Mayer, J E , Schell, J and Willmitzer, L (1988) Differential expression of genes in potato tubers after wounding Proc Natl Acad Sci U S A 85 1136-1140
- Muller-Rober, B T , Koßmann, J , Hannah, L C , Willmitzer, L and Sonnewald, U (1990) One of two different ADP-glucose pyrophosphorylase genes from potato responds strongly to elevated levels of sucrose Mol Gen Genet 224 136-146
- Powell, W and Uhrig, H (1987) Anther culture of *Solanum* genotypes Plant Cell Tissue and Organ Culture 11 13-24
- Ritter, E , Gebhardt, C and Salamini, F (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents Genetics 125 645-654
- Ritter, E , Debener, T , Barone, A , Salamini, F and Gebhardt, C (1991) RFLP mapping on potato chromosomes of two genes controlling extreme resistance to potato virus X (PVX) Mol Gen Genet 227 81-85
- Rosahl, S , Schmidt, R , Schell, J and Willmitzer, L (1986) Isolation and characterization of a gene from *Solanum tuberosum* encoding patatin, the major storage protein of potato tubers Mol Gen Genet 203 214-220
- Salaman, R (1985) The History and Social Influence of the Potato, revised impression (Hawkes, J G , ed) Cambridge University Press, Cambridge, New York, New Rochelle, Melbourne, Sydney
- Slocum, M K , Figdore, S S , Kennard, W C , Suzuki, J Y and Osborn, T C (1990) Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea* Theor Appl Genet 80 57-67
- Taylor, J L , Fritzemeier, K H , Hauser, I , Kombrink, E , Rohwer, F , Schroder, M , Strittmatter, G and Hahlbrock, K (1990) Structural analysis and activation by fungal infection of a gene

- encoding a pathogenesis-related protein in potato. *Mol. Plant-Microbe Interactions* 3: 72–77.
- Werr, W., Frommer, W.B., Maas, C. and Starlinger, P. (1985) Structure of the sucrose synthase gene on chromosome 9 of *Zea mays* L. *EMBO J.* 4: 1373–1380.
- Wolter, F.P., Fritz, C.C., Willmitzer, L. and Schell, J. (1988) *rbcS* genes in *Solanum tuberosum*: conservation of transit peptide and exon shuffling during evolution. *Proc. Natl. Acad. Sci. U.S.A.* 85: 846–850.

17. Rice molecular map

S.D. TANKSLEY

Department of Plant Breeding and Biometry, Cornell University, Ithaca, NY 14853-1901, U.S.A.

Contents

1. Introduction	286	4. Future work	287
2. History of genetic mapping in rice	286	5. References	290
3. RFLP map	287		

1. Introduction

Rice is ideally suited to genetic and molecular studies. Some of its positive features include a small nuclear genome (450 Mb), diploidy ($2n = 2x = 24$) and the ability to be transformed by exogenous DNA (Uchimaya et al. 1986; Arumuganathan and Earle 1991). Moreover, because of its importance as a crop species, scientists have been identifying and studying the genetic basis of morphological and physiological mutants for nearly a century.

2. History of genetic mapping in rice

As with most plant species, the first genetic linkage maps of rice were constructed with morphological markers (Kinoshita 1986 and references therein). By the mid 1980's isozymes were being added to the genetic maps of rice (Ranjhan et al. 1988; Wu et al. 1988). Mapping of both morphological and isozyme markers was greatly aided by the development of a series of primary trisomics which could be used to associate genetic loci (and thus linkage groups) with each of the 12 chromosomes (Khush and Singh 1986). In 1988, the first rice RFLP linkage map was reported (McCouch et al. 1988). This map was developed at Cornell University and contained 135 loci corresponding to single or low copy *Pst*I-generated genomic clones. Since that time, a large number of additional RFLP loci have been mapped in rice. The main RFLP mapping efforts on rice have been at Cornell University (sponsored by the Rockefeller Foundation) and at the Ministry of Fishery and Agriculture in Japan

(Kishimoto et al. 1989). RFLP maps developed by these two groups are now being integrated and together comprise more than 1000 loci corresponding to both genomic and cDNA clones. The map developed at Cornell currently contains approximately 600 loci and is the one presented in this chapter.

3. RFLP map

The RFLP map presented in Fig. 1 was based on a backcross population derived from the interspecific cross *Oryza sativa* × *O. longistaminata*. *O. longistaminata* is a wild rice from Africa which possesses the same genome (AA) as *Oryza sativa*. The high level of restriction fragment length polymorphism between the two species allows for great efficiency in genetic mapping. A total of 1222 map units was measured in this cross; however, the overall rate of recombination appears to be approximately 70% of that found in intraspecific crosses (Causse et al., in prep.). Total map units in intraspecific crosses would thus be expected to be approximately 1750 cM. This corresponds to a DNA: cM ratio of approximately 250 kb/cM which is significantly less than most crop species and very close to the value for *Arabidopsis thaliana*. This feature, combined with the ability to transform rice, makes this crop a likely target for the use of map-based cloning to isolate genes that have been localized on the RFLP map.

A number of morphological markers have also been mapped relative to RFLP markers and are shown on the map in Fig. 1. Of highest priority for rice breeders has been the mapping of genes important in rice production – especially disease and insect resistance genes. Rice blast (caused by the fungal pathogen *Pyricularia oryzae*) and bacterial leaf blight (caused by *Xanthomonas oryzae* pv. *oryzae*) represent two of the most serious disease problems in rice worldwide and several major genes conferring resistance to these pathogens have been mapped with respect to RFLPs (Yu et al. 1991; Ronald et al. 1992; McCouch, pers. comm.). Genes for photoperiod sensitivity and rice grain aroma have also been localized on the RFLP map (Mackill et al. 1992; Ahn et al. 1992).

4. Future work

Currently, a number of studies are in progress to use the rice RFLP map to identify quantitative trait loci (QTL) important in rice breeding. Some of the primary targets are traits that limit rice production in areas of high pathogen pressure or which experience climatic or edaphic extremes. High priorities include identification of QTLs for drought and salt tolerance and horizontal (or multigenic) disease resistance (McCouch, pers. comm.). In addition, studies are currently underway to determine the genetic basis of intersubspecific hybrid vigor (e.g. *indica* × *japonica*) (Xiao, pers. comm.). Heterotic yield increases in intersubspecific crosses are often double that found in intra-

5. References

- Ahn, N , Bollich, C N and Tanksley, S D (1992) RFLP tagging of a gene for aroma in rice Theor Appl Genet 84 825–828
- Khush, G S , Singh, R J , Sur, S C and Librojo, A L (1984) Primary trisomics of rice origin, morphology, cytology, and use in linkage mapping Genetics 107 141–163
- Kinoshita, T (1986) Standardization of gene symbols and linkage maps in rice In Rice Genetics, pp 215–228 Proc Int Rice Gen Symp , May 1985, IRRI, Los Banos, Philippines
- Kishimoto, N , Yamo, M , Tanaka, T , Saito, K , Nagamine, T , Kuhara, S , Kawase, M , Katsuta, M , Yoshimura, S , Yoshimura, A , Nakagahra, M and Iwata, N (1989) Linkage mapping of RFLP markers of rice nuclear DNA, morphological markers, and isozyme loci in rice (*Oryza sativa* L) Proc 6th Int Congr of SABRAO 489–498
- Mackill, D J , Salam, M A , Wang, Z Y and Tanksley, S D (1993) A major photoperiod-sensitivity gene tagged with RFLP and isozyme markers in rice Theor Appl Genet 85 536–540
- McCouch, S R , Kochert, G , Yu, Z H , Khush, G S , Coffman, W R and Tanksley, S D (1988) Molecular mapping of rice chromosomes Theor Appl Genet 76 815–829
- Ranjhan, S , Glaszmann, J L , Ramirez, D A and Khush, G S (1988) Chromosomal location of four isozyme loci by trisomic analysis in rice (*Oryza sativa* L) Theor Appl Genet 75 541–545
- Ronald, P C , Albano, B , Tabien, R , Abenes, L , Wu, K S , McCouch, S and Tanksley, S D (1993) Genetic and physical mapping of the rice bacterial blight disease resistance locus, *Xa21* Mol Gen Genet 236(1) 113–120
- Wu, K S , Glaszmann, J C and Khush, G S (1988) Chromosome location of ten isozyme loci in rice (*Oryza sativa* L) through trisomic analysis Biochem Genet 26 311–328
- Yu, Z H , Mackill, D J , Bonman, J M and Tanksley, S D (1991) Tagging genes for blast resistance in rice via linkage to RFLP markers Theor Appl Genet 81 471–476

18. Generation of a genetic map for *Sorghum bicolor*

JEFFREY L. BENNETZEN and ADMASU MELAKE-BERHAN

Department of Biological Sciences, Purdue University, W. Lafayette, IN 47907, U.S.A.

Contents

1. Introduction	291	4. Note added in proof	297
2. RFLP maps in sorghum	292	5. Acknowledgements	297
3. Future prospects	296	6. References	297

1. Introduction

Sorghum (*Sorghum bicolor* L. Moench) is one of the five most important grain crops worldwide, comprising the staple food source for over 300 million people in Africa and Asia (Doggett 1988). Sorghum is a very hardy plant that produces reasonable yields in locations with poor soils or limiting rainfall. In favorable environments, sorghum is a high yielding plant that has become a major feed crop throughout the developed and developing world. Sorghum is the closest relative to maize (*Zea mays*) and sugarcane (*Saccharum* species) among the major crops (Springer et al. 1989) and apparently arose as a unique species in Northeast Africa (Mann et al. 1983).

Over two hundred morphological traits have been identified in sorghum, many of them with characteristics quite similar to known allelic variants in maize (Schertz and Stephens 1966; Doggett 1988). Sorghum geneticists have also developed many useful tools, including a set of trisomics and both cytoplasmic and nuclear male sterility. Despite the importance of sorghum and the genetic tools available, the classical sorghum genetic map is woefully deficient. Only five multigene linkage groups have been identified and the most extensive group contains a mere ten linked loci (Doggett 1988). Hence, sorghum is a prime candidate for the generation of a genetic map based on restriction fragment length polymorphism (RFLP) technology.

Our laboratory initiated an RFLP analysis of sorghum in 1988 using cloned maize DNA as the source of probes. We found that maize low copy number sequences hybridized efficiently to sorghum, and detected polymorphisms frequently enough to generate a preliminary RFLP map (Hulbert et al. 1990).

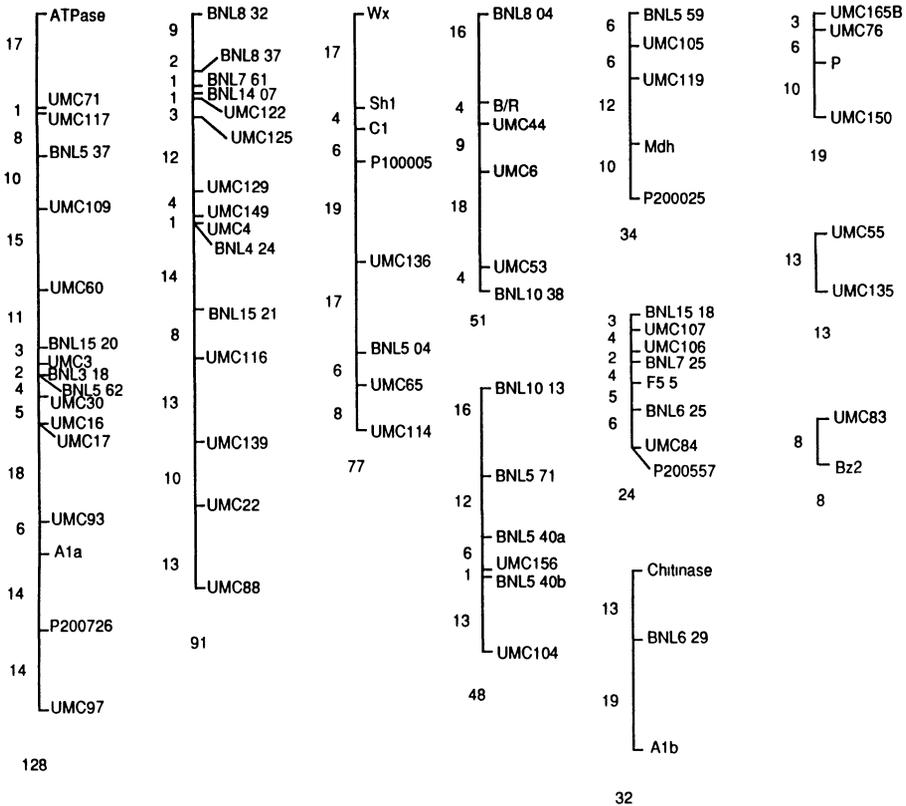
Since that time, we have expanded this map and several other laboratories have undertaken similar mapping studies.

2. RFLP maps in sorghum

For our current sorghum RFLP map, we have employed 250 maize genes and RFLP probes. All but 4 of these yielded one or more strong bands upon gel blot hybridization to total sorghum DNA. The hybridization conditions employed are those standardly used for high stringency, within species experiments (Hulbert et al. 1990). Band intensities generally ranged from about 10% to > 100% as strong as hybridization of the same probe to an adjacent lane containing maize DNA (Hulbert et al. 1990). The high intensities of these hybridizations are at least partly due to the three to four fold smaller genome size of sorghum relative to maize (Laurie and Bennett 1985; Michaelson et al. 1991); an agarose gel lane with four micrograms of sorghum DNA has 3 to 4 times as many genome equivalents as does an adjacent lane with four micrograms of maize DNA.

Of the 246 maize probes that hybridized well to sorghum DNA, 115 exhibited polymorphism between our two mapping parents. All but 11 of these were easily scored on DNAs derived from an 'immortal' F₃ representing a population of 55 F₂ progeny. These 104 polymorphic markers have been analyzed by neural manipulation of data compiled on the LINKAGE-I program (Suiter et al. 1983). In addition, we have mapped an RFLP derived from a cloned sorghum *Mdh* gene (Cretin et al. 1990). With the population size employed, we have chosen a recombination value of 20% or less as significant. Through this analysis, we were able to sort 77 markers into 11 linkage groups, some with only two or three linked RFLPs (Fig. 1). The largest of these contains 17 RFLP markers and covers about 128 cM. Overall, our linked RFLP markers cover about 525 cM. However, 28 markers were not linked to any other RFLP probe. Given the level of linkage significance that we have assigned, this indicates that the sorghum genetic map must contain a minimum of 1200 cM. We have collaborated with J.S. Neck and K. Schertz (Texas A&M University) on the integration of three isozyme markers into this map (data not shown). In addition, we have placed on this map single genes that determine seed color, plant color, and the presence of awns (data not shown). Further experiments are underway to increase the F₂ progeny analyzed to 71, to incorporate several additional morphological and isozyme traits, and to add many more maize and sorghum DNA markers to the developing RFLP map.

The parents chosen for our RFLP program were two sorghum lines representing different sorghum races from very different geographic locations. One parent is the caudatum line M91051, a zera zera cultivar from East Africa, and the other is Shanqui Red, a kaoliang of race bicolor from North Central China. The F₁ and subsequent progeny of this cross are quite vigorous and fully fertile. We have seen no significant segregation distortion for any of the



128

32

Fig 1 A 77 marker restriction fragment length polymorphism (RFLP) map of *Sorghum bicolor*. Linkage groups are presented in descending order, from left to right, of the number of markers identified in each group. Twenty eight other segregating, but not significantly linked, RFLP markers are not presented. The numbers below each linkage group indicate its approximate size (in centiMorgans) and the numbers to the left of each linkage group indicate the approximate genetic distance (also in centiMorgans) between each RFLP marker. The probes designated BNL (Burr et al 1988) and UMC (Hoisington and Coe 1990) were kindly provided by D Hoisington. The probes designated P (Grant et al 1989) and NPI (Helentjaris et al 1988) were generously supplied by D Grant. One or more polymorphic bands hybridizing to specific cloned maize genes were also placed on the map: *AI* (O'Reilly et al 1985) and *C1* (Paz-Ares et al 1986) from H Saedler, *Bz2* (Theres et al 1987) and *Sh1* (Geiser et al 1982) from R Kunze and P Starlinger, *R* (Ludwig et al 1989) and *Wx* (Shure et al 1983) from S Wessler, *B* from V Chandler (Chandler et al 1989), and *P* from T Peterson (Lechelt et al 1989). The *Mdh* gene was cloned from sorghum and was generously provided by C Cretin (Cretin et al 1990). The ATPase designation indicates a maize gene that we have cloned (Y-K Jin and J Bennetzen, unpub) which has striking homology to the plasma membrane proton ATPase gene of *Arabidopsis thaliana* (Harper et al 1989, Pardo and Serrano 1989). Chitinase refers to a maize chitinase gene cloned, and kindly provided to us, by A Kriz of the University of Illinois at Urbana (A Kriz, pers comm). The RFLP marker F5 5 is a fragment of DNA that we have cloned (P Springer and J Bennetzen, unpub) which flanked a *Mutator* transposable element and was located between the *Adh1* and *Tb* loci of maize.

RFLP probes that we have mapped. The degree of polymorphism between these two lines is quite high, about 50% of probes and about 30% of bands (Hulbert et al. 1990). In limited diversity studies with 25 maize RFLP probes on 35 additional sorghum lines broadly sampled from the world collection, we have found no other *Sorghum bicolor* subsp. *bicolor* lines that are significantly more different from each other than M91051 is from Shanqui Red (T. Richter and J. Bennetzen, unpub.).

As previously noted (Hulbert et al. 1990), our expanded data set has confirmed that the sorghum genome, like that of maize (Helentjaris et al. 1988), is duplicated for most RFLP probes. This is not at all surprising, since there are sorghum relatives (e.g., *Sorghum versicolor*) with five chromosome pairs compared to the ten chromosome pairs of *S. bicolor*. Therefore, it is likely that sorghum was derived from an $N = 5$ relative or relatives through auto- or allotetraploidization (Garber 1944). The DNA probes that give two or more bands in maize are the same probes that give two or more bands in sorghum. Hence, the duplications in the maize and sorghum genomes have enough similarities to suggest that this event or events occurred prior to the separate descent of these two species. Alternatively, these duplications could have been parallel or convergent phenomena with some as yet unknown level of selection for their specific nature.

These duplications make it particularly difficult to compare the sorghum and maize genetic maps; it is not always clear whether we are mapping the 'same' gene in maize and sorghum or are mapping different members of the same gene family in the two species. Primarily by using probes with a single band of hybridization in both species, we have been able to make some structural comparisons of the two genomes. We observe significant stretches of colinearity between the two species and also some evidence of chromosomal rearrangements (at least one major inversion and one major translocation) that differentiate the two species (Hulbert et al. 1990). Other regions have too few markers to make any solid conclusions at this time.

The generation of parallel genetic maps, using the same set of DNA probes for analysis of a closely related set of species, has distinct advantages for the analysis of chromosomal evolution. Our preliminary studies indicated that maize DNA probes could be used to map sugarcane and foxtail millet (*Setaria indica*) under the same conditions as those used for mapping sorghum, but that pearl millet (*Pennisetum glaucum*) and barley (*Hordeum vulgare*) did not have sufficient homology for this approach to work without using lower hybridization stringencies (Hulbert et al. 1990). We have tried a few genomic DNA probes from millet species as molecular markers in the characterization of sorghum. Two nuclear DNA probes from *Pennisetum* millets (kindly provided by P. Ozias-Akins, University of Georgia at Athens) were found to hybridize efficiently to maize DNA, but only one hybridized well to sorghum DNA under our high stringency conditions. At these same stringencies, we found that a nuclear DNA clone that we purified from finger millet (*Eleusine coracana*) hybridized very well to foxtail millet, and proso millet (*Panicum miliaceum*), but

only weakly to maize, sorghum, sugarcane, or barley (T. Richter and J. Bennetzen, unpub.). Similarly, a genomic DNA probe that we isolated from foxtail millet hybridized reasonably well to maize and sorghum DNAs, but did not hybridize well to finger millet, proso millet, sugarcane or barley DNAs (T. Richter and J. Bennetzen, unpub.). Hence, parallel genome organization studies comparing maize and sorghum with a wide range of millets, or barley and other species more distant from maize and sorghum, will probably require either multiple sources of probes or more permissive hybridization conditions.

At least four other academic programs are actively involved in the generation of genetic maps, based largely on RFLP technology, for *Sorghum bicolor*. Additional mapping studies and/or diversity analyses of sorghum are proposed or just underway in a number of other laboratories.

G. Hart and coworkers at Texas A&M University (Hart 1992) have begun their sorghum RFLP analyses with sorghum genomic *Pst*I DNA fragments and maize DNA probes derived from the maize RFLP programs at Brookhaven National Laboratory (Burr et al. 1988) and the University of Missouri at Columbia (Hoisington and Coe 1990). Based on diversity analyses of six sorghum lines, a cross between IS3620C (a guinea line) and BTx623 (an agronomic line derived from a cross between zera zera and kafir lines) was chosen for mapping. Tissues from 52 F₂ plants were originally used as the source of DNA for their studies, and an 'immortal' F₃ has been generated from these same plants. So far, 89 RFLP markers have been placed on this map and 17 linkage groups containing 70 markers have been identified. This map covers about 800 cM. Hart and associates plan to continue to RFLP map this population, and will also initiate studies with randomly amplified polymorphic DNAs (RAPDs) (Hart 1992).

R. Whitkus, J. Doebley (University of Minnesota), and M. Lee (Iowa State University) have employed 81 F₂ progeny of a cross between IS2482C (a bicolor race of *S. bicolor* subsp. *bicolor*) and A85213 (a virgatum accession of *S. bicolor* subsp. *verticilliflorum*) for the generation of a RFLP map (M. Lee, pers. comm.). Under their hybridization conditions, 79% of the maize RFLP probes employed gave a hybridization signal sufficient for analysis. Since the two parents were from different subspecies of *S. bicolor*, it is not surprising that most of the probes employed were found to be polymorphic. However, no evidence of lowered fertility or segregation distortion was observed in any generation derived from this cross (M. Lee, pers. comm.). Whitkus and coworkers' map contains 13 linkage groups with 85 maize RFLP probes and five isozyme traits covering about 950 cM. Their analyses predicted that about 30% of the sorghum genome is colinear with that of maize, and that colinear loci are separated by fewer centiMorgans on the average in sorghum than they are in maize (M. Lee, pers. comm.).

E. Ottaviano and coworkers at the University of Milan (Binelli et al. 1992) chose an F₂ population of 149 individuals derived from a cross between IS18729 (a caudatum-bicolor line) and IS24759 (a durra-caudatum) for their sorghum mapping study. Using maize DNA probes, they found that 154 out of

159 probes hybridized significantly well for RFLP analysis, and that 58 of these were polymorphic between their two parents. Thirty five of these probes were chosen for further analysis; those mapping to chromosomes 1, 3, 4, and 5 in maize. To date, 21 of these probes have been placed in five linkage groups and the 14 others were unlinked (Binelli et al. 1992). The genetic distance defined by the 21 linked markers in the five linkage groups was reported to be about 440 cM. However, since severe segregation distortion was observed within the progeny of this mapping cross (E. Ottaviano, pers. comm.), the map distances determined in this study may prove exceptional.

G. Ejeta and P. Goldsbrough at Purdue University have initiated a study involving the mapping of quantitative trait loci that determine preflowering and postflowering drought stress tolerance in sorghum (P. Goldsbrough, pers. comm.). Recombinant inbred progeny (Burr et al. 1988) of a cross between sorghum lines Tx7078 and B35 have been analyzed by use of RAPD markers. The RAPD approach was employed both for its speed and due to the fact that none of the six RFLP probes that they employed (chosen from maize DNA probes that were polymorphic and mapped in the Hulbert et al. (1990) study of sorghum) exhibited any polymorphism between their two parents. However, about 50% of the RAPD probes that they have used have detected useful variation between the two parental lines (P. Goldsbrough, pers. comm.).

3. Future prospects

Many laboratories have now begun investigations into the genomic organization of *Sorghum bicolor* and related species. Much of this work has been encouraged, supported, and coordinated by The Rockefeller Foundation. Hence, the sorghum mapping programs at the University of Milan, Texas A&M University, Iowa State University, and Purdue University routinely communicate and exchange their results. Other programs involved in sorghum RFLP mapping and/or diversity analysis are also participating in this informal network. Interactions between developed and developing country programs have been encouraged and initiated. This group plans to select a common set of well-spaced DNA probes to be mapped in all of our programs so that integration of the separate maps will be facilitated. Similarly, J. Glaszmann at the Institut de Recherches Agronomique Tropicales (Montpellier, France) was asked by this group to choose, and now has chosen, a set of sorghum lines that can be used as standards by those laboratories involved in characterization of sorghum diversity. Hence, the sorghum network (and an allied network of millet researchers headed by C. Liu, M. Gale (John Innes Centre, Norwich, U.K.) and J. Witcombe (University College of North Wales, Bangor, U.K.) hopes to use both the organizational and technical knowledge acquired by more advanced crop RFLP programs to facilitate similar studies in sorghum and related crop species.

Greatly improved genetic maps, particularly those derived from RFLP and

RAPD programs, can contribute immensely to future sorghum improvement by plant breeders. A number of laboratories are now placing important single gene and quantitative trait loci on the sorghum RFLP map. The next few years will see many important traits localized and also will see the use of RFLPs to enhance introduction of some of these traits into improved/adapted lines.

4. Note added in proof

Since this manuscript was submitted, Whitkus and coworkers (R. Whitkus, J. Doebley and M. Lee (1992) *Genetics* 132: 1119–1130) have published the most detailed sorghum genetic map to date, primarily using maize DNA probes as markers.

5. Acknowledgements

We express our gratitude to S. Hulbert for the initiation of these mapping studies at Purdue and for instructing us in the use of Linkage-1 to generate genetic maps. We also thank all of those individuals and organizations that generously supplied DNA probes and especially thank P. Goldsbrough, G. Hart, M. Lee and the late E. Ottaviano for providing unpublished information on their sorghum mapping programs. The research reported herein was generously supported by The Rockefeller Foundation.

6. References

- Binelli, G , Gianfranceschi, L , Pe M E , Taramino, G , Busso, C , Stenhouse, J and Ottaviano, E (1992) Similarity of maize and sorghum genomes as revealed by maize RFLP probes *Theor Appl Genet* 84 10–16
- Burr, B , Burr, F A , Thompson, K A , Albertsen, M C and Stuber, C W (1988) Gene mapping with recombinant inbreds in maize *Genetics* 118 519–526
- Chandler, V L , Radicella, J P , Robbins, T P , Chen, J and Turks, D (1989) Two regulatory genes of the maize anthocyanin pathway are homologous isolation of *B* utilizing *R* genomic sequences *Plant Cell* 1 1175–1183
- Cretin, C , Luchetta, P , Joly, C , Decottignies, P , Lepimiec, L , Gadal, P , Sallantin, M , Huet, J C and Pernollet, J C (1990) Primary structure of sorghum malate dehydrogenase (NADP) deduced from cDNA sequence homology with malate dehydrogenases (NAD) *Eur J Biochem* 192 299–303
- Doggett, H (1988) *Sorghum* Longman Singapore Publ Ltd , Singapore
- Garber, E D (1944) A cytological study of the genus *Sorghum* subsections *Parasorghum* and *Eusorghum* *Am Nat* 78 89–94
- Geiser, M , Weck, E , Doring, H P , Werr, W , Courage-Tebbe, U , Tillman, E and Starlinger, P (1982) Genomic clones of a wild-type allele and a transposable element-induced mutant allele of the sucrose synthase gene of *Zea mays* L *EMBO J* 1 1455–1460
- Grant, D , Nelson, W , Blair, D , Katt, M L , Martich, J , Smith, J S C , Bowen, S L , Tenborg, R A , Fincher, R R , Meier, R , Beavis, W D and Smith, O S (1989) Identification and

- localization of loci affecting yield and other agronomic traits in maize In T Helentjaris and B Burr (eds), Development and Application of Molecular Markers to Problems in Plant Genetics, pp 135-138 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Harper, J F , Surowy, T K and Sussman, M R (1989) Molecular cloning and sequence of cDNA encoding the plasma membrane proton pump (H^+ -ATPase) of *Arabidopsis thaliana* Proc Natl Acad Sci U S A 86 1234-1238
- Hart, G E (1992) Molecular mapping of the sorghum genome status and potential contributions to plant breeding In D Wilkinson (ed), Proc 46th Annual Corn and Sorghum Industry Res Conf American Seed Trade Assoc , pp 91-103 ASTA, Washington, DC
- Helentjaris, T , Weber, D L and Wright, S (1988) Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms Genetics 118 353-363
- Hoisington, D A and Coe, E H Jr (1990) Mapping in maize using RFLPs In J P Gustafson (ed), Gene Manipulation in Plant Improvement II, pp 331-352 Plenum Press, New York
- Hulbert, S H , Richter, T E , Axtell, J D and Bennetzen, J L (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes Proc Natl Acad Sci U S A 87 4251-4255
- Laurie, D A and Bennett, M D (1985) Nuclear DNA content in the genera *Zea* and *Sorghum* Intergeneric, interspecific, and intraspecific variation Heredity 55 307-313
- Lechelt, C , Peterson, T , Laird, A , Chen, J , Dellaporta, S L , Dennis, E , Peacock, W J and Starlinger, P (1989) Isolation and molecular analysis of the maize P locus Mol Gen Genet 219 225-234
- Ludwig, S R , Habera, L F , Dellaporta, S L and Wessler, S R (1989) *Lc*, a member of the maize *R* gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region Proc Natl Acad Sci U S A 86 7092-7096
- Mann, J A , Kimber, C T and Miller, F R (1983) The origin and early cultivation of sorghums in Africa Bulletin no 54, Texas A&M University, College Station, TX
- Michaelson, M J , Price, H J , Ellison, J R and Johnston, J S (1991) Comparison of plant DNA contents determined by Feulgen microspectrophotometry and laser flow cytometry Am J Bot 78 183-188
- O'Reilly, C , Shepherd, N S , Pereira, A , Schwarz-Sommer, Z , Bertram, I , Robertson, D S , Peterson, P A and Saedler, H (1985) Molecular cloning of the *al* locus of *Zea mays* using the transposable elements *En* and *Mul* EMBO J 4 877-882
- Pardo, J M and Serrano R (1989) Structure of a plasma membrane H^+ -ATPase gene from the plant *Arabidopsis thaliana* J Biol Chem 264 8557-8562
- Paz-Ares, J , Wienand, U , Peterson, P A and Saedler, H (1986) Molecular cloning of the *C* locus of *Zea mays* a locus regulating the anthocyanin pathway EMBO J 5 829-833
- Schertz, K F and Stephens, J C (1966) Compilation of gene symbols recommended revision and summary of linkages for inherited characters of *Sorghum vulgare* Pers Texas A&M University Technical Monograph, College Station, TX
- Springer, P S , Zimmer, E A and Bennetzen, J L (1989) Genomic organization of the ribosomal DNA of sorghum and its close relatives Theor Appl Genet 77 844-850
- Shure, M , Wessler, S and Fedoroff, N (1983) Molecular identification and isolation of the *waxy* locus in maize Cell 35 225-233
- Suiter, K A , Wendel, J F and Case, J S (1983) Linkage-1 a PASCAL computer program for the detection and analysis of genetic linkages J Hered 74 203-204
- Theres, N , Scheele, T and Starlinger, P (1987) Cloning of the *Bz2* locus of *Zea mays* using the transposable element *Ds* as a gene tag Mol Gen Genet 209 193-197

19. RFLP map of soybean

RANDY C. SHOEMAKER

USDA-ARS-FCR, Department of Agronomy, Iowa State University, Ames, IA 50011, U.S.A.

Contents

1. Historical perspective	299	5. Genome duplication	304
2. Probe and map development	300	6. Validity of an interspecific genetic map	307
3. Integrating classical and molecular markers	301	7. References	307
4. Quantitative trait loci	302		

1. Historical perspective

The soybean is one of the oldest cultivated crops. It first emerged as a domesticated plant around the 11th century B.C. (Hymowitz 1970), although references to soybean appear in books written over 4500 years ago (Smith and Huysen 1987). The soybean was first planted in the United States in Thunderbolt, GA, in 1765 (Hymowitz and Harlan 1983). It was originally planted as a forage crop but by the early 1900's became of interest as an oilseed crop (Smith and Huysen 1987), an interest that continues today.

Soybean production was initiated in this country through the use of cultivars introduced from other countries. These plant introductions were evaluated and the most desirable ones were released for commercial use (Fehr 1987). During the 1940's cultivars were selected from crosses made between plant introductions. Often, the superior cultivars selected from this initial round of hybridizations were used as parents in the next cycle of hybridizations. This is a breeding scheme that continues today (Fehr 1987).

A limited number of accessions were used to derive the cultivars commonly grown in the U.S. today. An analysis of the pedigrees of cultivars in the Northern germplasm collection indicates that 88% of their collective genome was derived from just 10 accessions (Delannay et al. 1983; Specht and Williams 1984). Within the Southern germplasm the situation is even more extreme. Seventy percent of the genome of commonly grown Southern cultivars was shown to be contributed by as few as seven accessions (Delannay et al. 1983).

This history of breeding and cultivar development in soybean tends to limit the amount of genetic diversity found among elite lines.

2. Probe and map development

The soybean genome contains an estimated 1.29×10^9 bp (Gurley et al. 1979) to 1.81×10^9 bp (Goldberg 1978) for 1n DNA content. This genome is approximately 40–60% repetitive sequences (Gurley et al. 1979; Goldberg 1978). The majority (65–70%) of single-copy sequences have a short period interspersion with single-copy sequences of 1.1–1.4 kb alternating with repetitive sequence elements of 0.3–0.4 kb (Gurley et al. 1979).

Analysis of pachytene chromosomes has shown that over 35% of the soybean genome is made up of heterochromatin – the short arms of six of the 20 bivalents are completely heterochromatic (Singh and Hymowitz 1988).

Low- or single-copy clones are required in establishing an RFLP map. An initial effort to create a soybean RFLP map involved the development of a random genomic library digested with *Sau3AI* and cloned into M13 digested with *BamHI* (Apuya et al. 1988). To differentiate single-copy clones from clones carrying repetitive DNA, the clones were probed with radioactive total soybean genomic DNA. Single- and low-copy clones were identified by their lack of hybridization (Apuya et al. 1988). This approach is effective but labor intensive.

A high percentage of plant DNA is modified by methylation of the cytosine base. Many restriction endonucleases will not cleave DNA containing methylated cytosines within certain sequences. Keim and Shoemaker (1988) used the restriction endonuclease *PstI*, a methylation-sensitive enzyme to construct a genomic library. Sequences in and adjacent to transcribed regions generally are unmethylated and therefore will be cleaved by this enzyme (Burr et al. 1988). Using this technique, a library of recombinant DNA was developed that is primarily single-copy DNA sequence (Keim and Shoemaker 1988).

Total DNA was isolated from leaves of soybean seedlings (Keim et al. 1988). This DNA was digested with a 20-fold excess of *PstI*. DNA fragments were electrophoretically separated in low-melting agarose. Fragments in the estimated size range of 0.5–3.0 kb were extracted from the gel, and phenol extracted and cloned into the plasmid vector, pBS+. Recombinant molecules were transformed into the *E. coli* strain DH5 (α). Approximately 80–85% of the clones represent single- or low-copy DNA sequence (Keim and Shoemaker 1988; Keim et al. 1990a).

When tested against DNA from the *G. max* breeding line A81-356022 and *G. soja* accession PI 468.916 digested with restriction enzymes *DraI*, *TaqI*, *HindIII*, *EcoRI*, or *EcoRV*, approximately 40% of the random genomic probes detect polymorphisms. About half of the polymorphic probes detect polymorphisms with two or more enzymes, suggesting that DNA rearrangements are the cause of the variability (Keim et al. 1990). This was substantiated through RFLP mapping of polymorphic regions (Apuya et al. 1988). Approximately 10% of the

probes detect 'dominant' markers, i.e., the heterozygote classes are indistinguishable from dominant classes. These 'dominant' markers could be the result of additions or deletions within one of the genotypes.

The map was generated by analysis of marker segregation among 60 F_2 individuals. The computer program 'MAPMAKER' (Lander et al. 1987) was used to create the best loci order. A minimum LOD score of 3.0 was used for the pairwise linkage analysis in all instances except two (pA-203 and pT-153b), where a LOD score of 2.8 was accepted. Three-hundred fifty-nine qualitative markers have been placed on this map. These markers include random genomic clones, clones of known genes, isoenzyme loci, and a variety of morphological and developmental classical markers. This map includes 20 linkage groups of three or more markers and four linkage groups containing only two-point linkages. The linkage map encompasses approximately 2900 cM.

3. Integrating classical and molecular markers

Historically, the construction of linkage maps comprised of conventional phenotypic genetic markers has constituted the traditional means of genome analysis. However, conventional map development in soybean has proceeded slowly, due primarily to the difficulty in performing crosses and generating large numbers of hybrid seed, the lack of detailed cytogenetic markers, and a paucity of genetic variation in the germplasm. Thus, compared to other economically important species, the resolution of the classical map is poor. Currently, the classical soybean genetic linkage map contains only 49 linked markers which covers approximately 530 map units (Palmer and Kiang 1990). Many other qualitative traits remain unmapped.

To fully exploit the potential of a molecular genetic map, it is necessary to integrate molecular and conventional markers into a unified linkage map (Table 3). This can be accomplished through painstaking segregation and linkage analysis in entire populations using both conventional and molecular markers; a technique used successfully by Landau-Ellis (1991) to map the locus controlling supernodulation (*nts*) in soybean. However, by screening near-isogenic lines (NIL), many molecular markers that are linked to the conventional marker can be efficiently identified at one time in donor and recurrent parents (Young et al. 1988).

Integrating classical markers into the molecular map is a potentially efficient process for soybeans, since their germplasm collection contains an extensive number of NILs (Bernard 1976). The potential of NILs in integrating soybean conventional and molecular linkage maps was discussed in depth by Muehlbauer et al. (1988). They calculated that if a BC_5S_1 NIL was screened with 100 randomly chosen loci, assuming polymorphisms existed between the recurrent parent and the donor parent at all loci, four loci should detect donor DNA and two or three of them could be expected to be genetically linked to the introgressed gene. Evaluating 63 NILs, each possessing an introgressed

conventional gene with 12 isozyme loci, 5 presumptive linkages were observed by Muehlbauer et al. (1989). Segregation analysis confirmed linkages of *Enp* (endopeptidase) with *ln* (narrow leaflet) ($9.38 \pm 1.55\%$) and *Mpi* (mannose-6-phosphate isomerase) with *dt2* (indeterminate stem) ($16.07 \pm 6.43\%$). Using 15 RFLP markers, Muehlbauer et al. (1991) screened 116 NILs. Fifteen polymorphisms were observed where the NIL possessed the donor parent allele. Segregation analysis confirmed linkages of pK-3 with *p*₁ (pubescent pubescence type) (16.2 ± 5.0), pK-3 with *r* (brown seed) (14.3 ± 4.6) and pK-472 with *Lf*₁ (5-foliolate) (14.10 ± 4.82) (Table 2).

A series of isolines containing one or two alleles for resistance to phytophthora root rot (*Rps1-6*) and the locus conferring ineffective nodulation (*Rj2*) were screened with 141 mapped RFLP markers (Diers et al. 1992b). At least one polymorphism was detected between each NIL and the recurrent parent. Segregation analysis of F₂:F₃ lines confirmed linkages of *Rps1* to pK-418 and pA-280; *Rps2* and *Rj2* to pA-233 and pA-199; *Rps3* to pA-186 and pR-45; *Rps4* to pA-586; and *Rps5* to pT-5. However, even though linkages have been established, the position of some of these markers relative to other mapped molecular markers remains ambiguous.

4. Quantitative trait loci

A limited number of studies have been conducted towards identifying quantitative trait loci (QTLs) in the soybean genome. Hard seededness, a quantitative trait that affects germination rate, viability, and quality of stored seeds, was evaluated using 60 F₂-derived lines and 72 markers (Keim et al. 1990b). Five genomic regions associated with seed hardness were identified. These regions explained 71% of the variation observed for seed hardness, and one region, associated with the *i* locus (conditioning seed coat color) accounted for 32% of the variation (Table 1) (Keim et al. 1990b). Also using 60 F₂-derived lines and 150 markers, Keim and co-workers (1990a) detected QTLs affecting the agronomically important traits of maturity (R1 and R8), seed-fill, stem diameter, stem length, canopy height, leaf width, and leaf length. The variation explained by each marker ranged from 16% to 24% (Table 1). Possible QTLs for iron deficiency in soybean were identified using 60 F₂-derived lines and 272 markers (Diers et al. 1991). Three markers were significantly associated with iron efficiency. Two of these markers were linked and explained 31% and 25% of the variation observed; the other marker explained 17% of the variation. However, upon re-establishing the crosses and repeating the experiment independently, none of the markers were found to be significant.

Soybean is grown primarily for the protein and oil processed from its seed (Smith and Huyser 1987). Both protein and oil content are quantitatively inherited (Burton 1985; Wilcox 1985). Two hundred fifty-two markers were used to evaluate replicated trials of 60 F₂-derived lines for variation in protein and oil (Diers et al. 1992a). Significant associations were found with eight

Table 1. Markers that have been associated with significant variation for quantitative traits. All variation shown is significant at probabilities less than 0.01

Trait	R ² value	Loci	Linkage group	Reference
Hard seededness	0.32	<i>i</i>	A	Keim et al. (1990b)
	0.34	T-153a	A	"
Seed protein	0.42	K-11a	I	Diers et al. (1992a)
	0.39	A-407a	I	"
	0.24	A-144	I	"
	0.25	A-688	I	"
	0.24	SAC-7a	E	"
	0.19	A-242b	E	"
	0.16	A-23	L	"
	0.12	A-245a	H	"
Seed oil	0.43	SAC-7a	E	Diers et al. (1992a)
	0.39	A-242b	E	"
	0.32	A-23	L	"
	0.27	<i>pb</i>	E	"
	0.27	K-11a	I	"
	0.28	A-407a	I	"
	0.23	A-454	E	"
	0.22	K-229	E	"
	0.18	A-203	E	"
	Leaf width	0.24	A-111	A
0.17		K-390	A	"
0.16		K-411	D	"
Leaf length	0.18	R-13b	D	Keim et al. (1990a)
	0.19	K-478a	D	"
Stem diameter	0.24	G-17.3a	L	Keim et al. (1990a)
	0.17	K-385	L	"
	0.17	R-201	L	"
Canopy height	0.16	K-390	A	Keim et al. (1990a)
	0.20	R-13b	D	"
First flower (R1)	0.21	K-365	C	Keim et al. (1990a)
	0.23	K-474a	C	"
	0.21	K-474b	C	"
Pod maturity (R8)	0.18	K-472	C	Keim et al. (1990a)
	0.18	R-13b	D	"
	0.17	K-365	C	"
	0.18	K-474a	C	"
	0.21	K-474b	C	"
Seedfill(R1 to R8)	0.18	G-8.15	J	Keim et al. (1990a)

markers for protein and nine markers for oil (Table 1). Because of the inverse relationship between protein and oil content (Burton 1985), most of the markers significant for one trait were also significant for the other.

Table 2. Probes detecting sequences genetically linked to qualitatively inherited genes/traits. Because of the possibility of duplicate loci the precise map position of these genes/traits remains ambiguous in some cases

Sequence gene/trait	Linked probes	Description	References
P1	pK-3	glabrous pubescence	Muehlbauer et al. (1991)
r	pK-3	seed coat color brown	„
Lf1	pK-472	pentafoliate leaf	„
nts	pA-132	increased nitrogen fixation in presence of nitrate	Landau-Ellis et al. 1991
Rps1	pA-36	phytophthora root rot resistance	Diers et al. (1992b)
	pK-418		
	pA-71		
Rps2	pA-280	phytophthora root rot resistance	„
	pA-233		
Rps3	pA-199b	phytophthora root rot resistance	„
	pR-45		
Rps4	pA-186	phytophthora root rot resistance	„
	pA-586		
Rps5	pT-5	phytophthora root rot resistance	„
Rj2	pA-233	ineffective nodulation	„
	pA-199b		

Table 3. Genes/traits that have been assigned map positions

Gene/trait	Description	Linkage group	Reference
pb	pubescent tip	E	Keim et al. (1990)
i	seed coat color	A	„
SAc3	actin gene	A	„
SAc7a	actin gene	E	Diers et al. (1992)
SAc7b	actin gene	A	„
B	seed coat luster	F	„
Dia	diaphorase	C	„
Ap	acid phosphatase	A	„
Me	malic enzyme	D	„
Vsp27	vegetative storage protein	A	„

5. Genome duplication

Most genera of the Phaseoleae have a genome complement of $2n = 22$. Lackey (1980) suggested that *Glycine* was probably derived from a diploid ancestor ($n = 11$) which underwent an aneuploid loss to $n = 10$ and subsequent polyploidation to yield the present $2n = 2x = 40$ genome size (Palmer and Kilen 1987). Because soybean behaves like a diploid we regard $2n = 40$ as the diploid chromosome number.

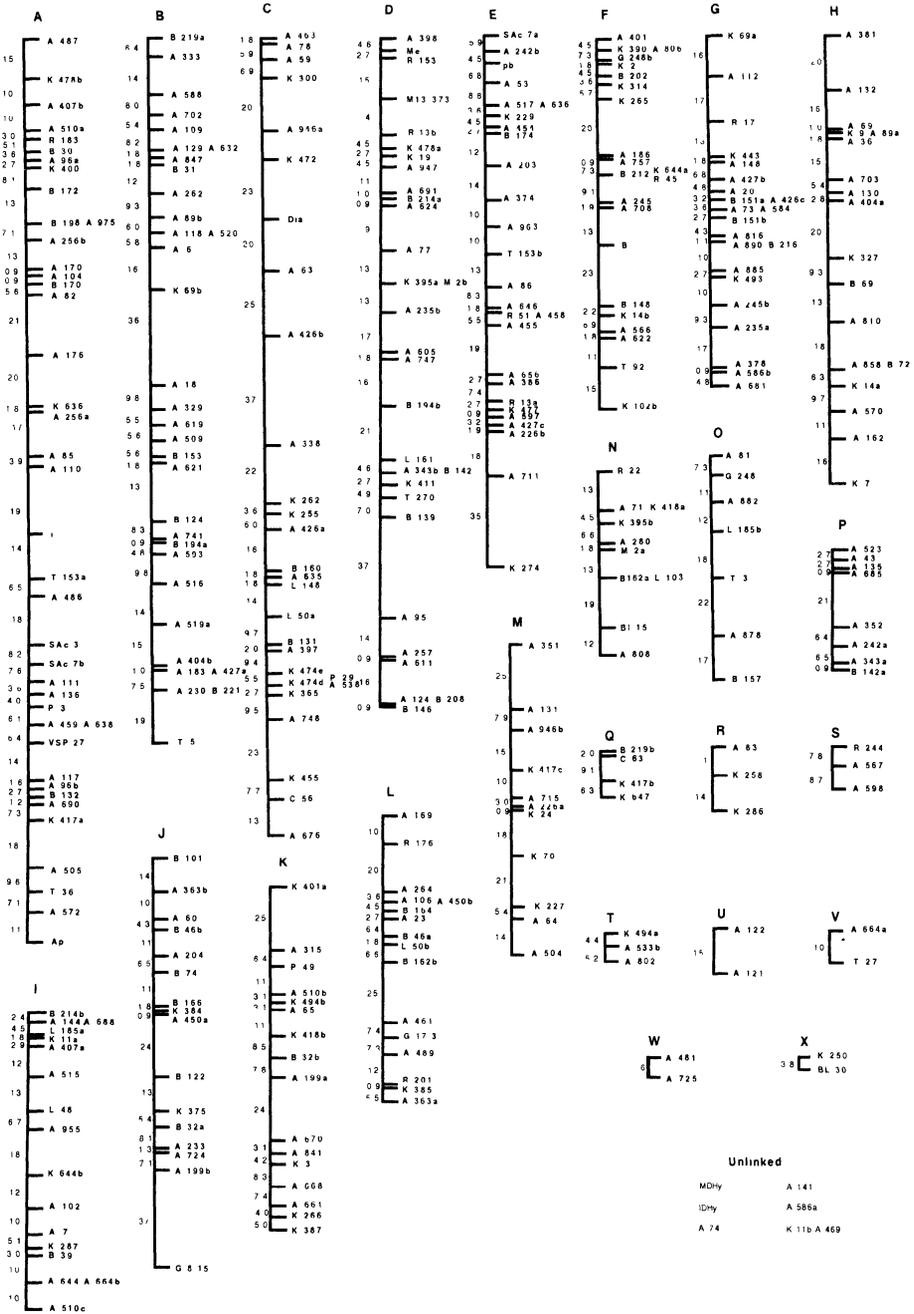
Genetic evidence of gene duplication suggests that soybean is a polyploid (Buttery and Buzzell 1976; Palmer and Kilen 1987). Hymowitz and Singh (1987) have suggested that the soybean be regarded as a stable tetraploid with diploidized genomes. Soybean possesses many examples of qualitative traits controlled by two loci (Zobel 1983; Palmer and Kilen 1987). Three-hundred fifty-nine loci have been mapped in our interspecific cross using standard recombinant DNA clones of soybean genomic DNA as probes. Forty (11%) of these probes detected independent, duplicate loci, while four (0.8%) detected independent, triplicate loci.

There is a strong tendency for polyploids to evolve into a diploid state through sequence divergence and chromosome rearrangement (Leipold and Schmidtke 1982). This type of genome rearrangement was observed during the analysis and comparison of the leghemoglobin genes of *Phaseolus vulgaris* and *Glycine max* (Lee and Verma 1984). Also, studies of ribosomal RNA (rRNA) gene sequences in soybean have indicated the rRNA gene sites have been eliminated during diploidization (Skorupska et al. 1989).

If soybean were to be a relatively new polyploid, or if the genome had remained relatively stable after polyploidation, distinct homoeologous chromosomes would be evident upon an analysis of linkage associations of duplicate molecular loci. In only a few instances did a linked set of duplicate loci seem to delineate a homoeologous chromosomal segment. For example, linkage group 'J' contains five markers associated with duplicate loci (Fig. 1). Two of those markers (pB-32a and pA-199b) are linked at approximately 16 map units. Their duplicate loci (pB-32a and pA-199a) also are linked at approximately 8 map units on linkage group 'K'. This suggests that a homoeologous linkage block has been retained. Additionally, three other markers (pA-363b, pB-46b and pA-450a) from linkage group 'J' exist in a similar linear order on linkage group 'L'. However, the distance between markers is not proportional between linkage groups. The distance between pB-46b and pA-363b is 14 map units on linkage group 'J', while pB-46a and pA-363a are separated by 66 map units on linkage group 'L'. The grossly disproportionate recombination distances may be explained by a disparity between physical distance and recombination distance or could have arisen if, during the evolution of the soybean genome, additional DNA sequences had been inserted between pB-46a and pA-363a. Our mapping data indicate that these sequences may have homoeologous sequences on linkage groups 'C' and 'N'.

It is important to remember that failure to detect duplicate loci does not preclude their existence. Most of our probes do indeed hybridize to multiple fragments, but, duplicate loci can only be detected if a polymorphism is present between parental genotypes and map at different sites. Therefore, sequences may exist between pB-46b and pA-363b on linkage group 'J' also, but simply remain undetected.

In most instances duplicate loci within one linkage group were not found together in another distinct homoeologous group, but instead seemed to be scattered randomly throughout the genome. An example of apparent



'scrambling' of the genome can be seen on linkage group 'E', where six markers detect duplicate, or in one case, triplicate loci. These six markers are duplicated in six separate linkage groups. It is difficult to reconcile the arrangement of homoeologous segments with the arrangement expected from a tetraploid (Helentjaris et al 1988). However, the independent translocation events necessary to account for these arrangements could be expected during the 'diploidization' of the ancient polyploid soybean (Zobel 1983). Studies of rRNA gene sequences in soybean have indicated that the rRNA gene sites have been eliminated during diploidization (Skorupska et al 1989).

Because soybean probes detect multiple fragments in a Southern hybridization experiment, and because each fragment could represent a different locus on a chromosome, it is very important to map each polymorphism. Without this information, the map location of polymorphic fragments must remain ambiguous.

6. Validity of an interspecific genetic map

The use of interspecific crosses in the construction of genetic maps creates the possibility of map aberrations resulting from cytogenetic variations such as inversions or translocations. However, pachytene chromosome analysis suggests that *G. max* and *G. soja* carry similar genomes. This genome is designated 'GG' (Singh and Hymowitz 1988).

Chromosome aberrations have been found in *G. max* at a low frequency, but occur at a higher frequency in *G. soja*. Forty-six of fifty-six *G. soja* accessions from China and the Soviet Union were determined to contain chromosome interchanges (Palmer et al 1987). PI 468 916, the *G. soja* accession used in the development of our mapping population, did not contain a chromosome aberration relative to *G. max* (Palmer et al 1987). Therefore, except for possible minor recombinational differences, the genetic map constructed from a cross between A81-356022 and PI 468916 represents an accurate linear representation of the soybean genome.

7. References

- Apuya, N R , Frazier, B L , Keim, P , Roth, E J and Lark, K G (1988) Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merrill. *Theor Appl Genet* 75: 889-901.
- Bernard R L (1976) United States national germplasm collections. In L D Hill (ed) *World Soybean Research* pp 286-289. Interstate Printers and Publ, Danville, IL.
- Burr, B , Burr, F A , Thompson, K H , Albertsen, M C and Stuber C W (1988) Gene mapping with recombinant inbreds in maize. *Genetics* 118: 519-526.
- Burton, J W (1985) Breeding soybeans for improved protein quantity and quality. In R Shibles (ed), *Proc 3rd World Soybean Res Conf*, pp 361-367. Westview Press, Boulder, CO.
- Buttery, B R and Buzzell R I (1976) Flavonol glycoside genes and photosynthesis in soybeans. *Crop Sci* 16: 547-550.

- Delannay, X, Rodgers, D M and Palmer, R G (1983) Relative genetic contributions among ancestral lines to North American soybean cultivars *Crop Sci* 23 944-949
- Diers, B W, Keim, P, Fehr, W R and Shoemaker, R C (1992a) RFLP analysis of soybean seed protein and oil content *Theor Appl Genet* 83 608-612
- Diers, B W, Mansur, L, Imsande, J and Shoemaker, R C (1992b) Mapping of Phytophthora resistance loci in soybean with restriction fragment length polymorphism markers *Crop Sci* 32 377-383
- Diers, B W, Cianzio, S R and Shoemaker, R C (1991) Possible identification of quantitative trait loci affecting iron efficiency in soybean VIth Int Symp on Iron Nutrition and Interactions in Plants Utah State University, Logan, UT
- Fehr, W R (1987) Breeding methods for cultivar development In J R Wilcox (ed), Soybeans Improvement, Production, and Uses, second ed, No 16, pp 249-293 American Society of Agronomy, Inc, Crop Science Society of America, Inc, and Soil Science Society of America, Inc, Madison, WI
- Goldberg, R B (1978) DNA sequence organization in the soybean plant *Biochem Genet* 16 45-68
- Gurley, W B, Hepburn, A G and Key, J L (1979) Sequence organization of the soybean genome *Biochem Biophys Acta* 561 167-183
- Helentjaris, T, King, G, Slocum, M, Siedenstrang, C and Wegman, S (1985) Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding *Plant Mol Biol* 5 109-118
- Helentjaris, T, Weber, D and Wright, S (1988) Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms *Genetics* 118 353-363
- Hymowitz, T (1970) On the domestication of the soybean *Econ Bot* 24 408-421
- Hymowitz, T and Harlan, J R (1983) Introduction of soybean to North America by Samuel Bowen in 1765 *Econ Bot* 37 371-379
- Hymowitz, T and Singh, R J (1987) Taxonomy and speciation In J R Wilcox (ed), Soybeans Improvement, Production, and Uses, second ed, No 16, pp 23-48 American Society of Agronomy, Inc, Crop Science Society of America, Inc, and Soil Science Society of America, Inc, Madison, WI
- Hymowitz, T, Palmer, R G and Singh, R J (1991) Cytogenetics of the genus *Glycine* In T Tsuchiya and P K Gupta (eds), Chromosome Engineering in Plants Genetics, Breeding, Evolution, Part B, pp 53-63 Elsevier Science Publishers B V, Amsterdam
- Keim, P and Shoemaker, R C (1988) Construction of a random recombinant DNA library that is primarily single copy sequences *Soy Genet News* 15 147-148
- Keim, P, Olson, T C and Shoemaker, R C (1988) A rapid protocol for isolating soybean DNA *Soy Genet News* 15 150-152
- Keim, P, Shoemaker, R C and Palmer, R G (1989) Restriction fragment length polymorphism diversity in soybean *Theor Appl Genet* 77 786-792
- Keim, P, Diers, B W, Olson, T C and Shoemaker, R C (1990a) RFLP mapping in soybean association between marker loci and variation in quantitative traits *Genetics* 126 735-742
- Keim, P, Diers, B W and Shoemaker, R C (1990b) Genetic analysis of soybean hard seededness with molecular markers *Theor Appl Genet* 79 465-469
- Lacky, J A (1980) Chromosome numbers in the Phaseoleae (Fabaceae Faboideae) and their relation to taxonomy *Am J Bot* 67 595-602
- Landau-Ellis, D, Angermuller, S, Shoemaker, R C and Gresshoff, P M (1991) The genetic locus controlling supernodulation in soybean (*Glycine max* L) co-segregates tightly with a cloned molecular marker *Mol Gen Genet* 228 221-226
- Lander, E S, Green, P, Abrahamson, J, Barlow, A, Daly, M, Lincoln, S E and Newburg, L (1987) MAPMAKER An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations *Genetics* 1 174-181
- Lee, J S and Verma, D S (1984) Structure and chromosomal arrangement of leghemoglobin genes in kidney bean suggest divergence in soybean leghemoglobin gene loci following tetraploidization *EMBO J* 12 2745-2752

- Leipold, M and Schmidtke, J (1982) Gene expression in phylogenetically polyploid organisms In G Dover and R Flavell (eds), Genome Evolution, pp 219–236 Academic Press, New York
- Muehlbauer, G J , Specht, J E , Thomas-Compton, M A , Staswick, P E and Bernard, R L (1988) Near isogenic lines – a potential resource in the integration of conventional and molecular marker linkage maps Crop Sci 28 729–735
- Muehlbauer, G J , Specht, J E , Staswick, P E , Graef, G L and Thomas-Compton, M A (1989) Application of the near-isogenic line gene mapping technique to isozyme markers Crop Sci 29 1548–1553
- Muehlbauer, G J , Staswick, P E , Specht, J E , Graef, G L , Shoemaker, R C and Keim, P (1991) RFLP mapping using near-isogenic lines in the soybean [*Glycine max* (L) Merr] Theor Appl Genet 81 189–198
- Palmer, R G and Kiang, Y T (1990) Linkage map of soybean (*Glycine max* L Merr) In S J O'Brian (ed), Genetic Maps, pp 668–693 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Palmer, R G and Kilen, T C (1987) Qualitative Genetics and Cytogenetics In J R Wilcox (ed), Soybeans Improvement, Production, and Uses, second ed , No 16, pp 135–209 American Society of Agronomy, Inc , Crop Science Society of America, Inc , and Soil Science Society of America, Inc , Madison, WI
- Palmer, R G , Newhouse, K E , Graybosch, R A and Delannay, X (1987) Chromosome structure of the wild soybean Accessions from China and the Soviet Union of *Glycine soja* Sieb & Zucc J Hered 78 243–247
- Singh, R J and Hymowitz, T (1988) The genomic relationship between *Glycine max* (L) Merr and *G soja* Sieb and Zucc as revealed by pachytene chromosomal analysis Theor Appl Genet 76 705–711
- Skorupska, H , Albertsen, M C , Langholz, K D and Palmer, R G (1989) Detection of ribosomal RNA genes in soybean, *Glycine max* (L) Merr by *in situ* hybridization Genome 32 1091–1095
- Smith, K J and Huyser, W (1987) World distribution and significance of soybean In J R Wilcox (ed), Soybeans Improvement, Production, and Uses, second ed , No 16, pp 1–22 American Society of Agronomy, Inc , Crop Science Society of America, Inc , and Soil Science Society of America, Inc , Madison, WI
- Specht, J E and Williams, J H (1984) Contribution of genetic technology to soybean productivity retrospect and prospect In W F Fehr (ed), Genetic Contributions to Yield Gains of Five Major Crop Plants, pp 49–74 American Society of Agronomy, Madison, WI
- Wilcox, J R (1985) Breeding soybeans for improved oil quantity and quality In R Shibles (ed), Proc 3rd World Soybean Res Conf , pp 380–386 Westview Press, Boulder, CO
- Young, N D , Zamir, D , Ganai, M W and Tanksley, S D (1988) Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the *Tm-2a* gene in tomato Genetics 120 579–585
- Zobel, R W (1983) Genic duplication a significant constraint to molecular and cellular genetic manipulation in plants Comments Mol Cell Biophys 1 355–364

20. Tomato molecular map

S.D. TANKSLEY

Department of Plant Breeding and Biometry, 252 Emerson Hall, Cornell University, Ithaca, NY 14853, U.S.A.

Contents

1. Introduction	310	2.2. Heterogeneity of crossing over along the chromosomes	321
1.1. History of mapping in tomato	311	3. Future prospects	321
2. Current status of tomato molecular linkage map	315	4. References	323
2.1. Localization of telomeres	321		

1. Introduction

Tomato (*Lycopersicon esculentum* L.) is well studied among higher plants and has several characteristics, including diploidy ($2n = 2x = 24$), self pollination and a relatively short generation time, that make it ideal for genetic studies. During the past half-century, the cultivated tomato and its wild relatives have been the subject of numerous and in-depth studies encompassing systematics, transmission genetics and population genetics, cytogenetics and mutation research (for reviews see Rick 1975; Rick and Yoder 1988). It has a nuclear genome that contains approximately 950 Mb of DNA (Arumuganathan and Earle 1991). This is considerably larger than *Arabidopsis*, but relatively small among crop plants. For comparison, rice has a genome size of 450 Mb, maize, 2500 Mb and wheat, 16000 Mb (Arumuganathan and Earle 1991). Moreover, the tomato genome has a large portion of single copy DNA (approximately 80% of the DNA behaves as single copy at high stringency, Zamir and Tanksley 1988) – a beneficial attribute with respect to chromosome walking. The major repetitive elements have also been cloned and characterized (Ganal et al. 1988) and tomato is the only plant species for which there is detailed knowledge about the long-range structure of telomeres (Ganal et al. 1991).

1.1. History of mapping in tomato

Genetic mapping in tomato began in the early parts of this century and by mid century there was already a respectable linkage map available for this species (Butler 1951; Clayberg et al. 1966). Over the years, the map was expanded and refined and by 1975 more than 200 morphological markers had been assigned to the 12 linkage groups (Rick 1975). Concomitant with the development of a morphological linkage map was the development of a physical map based on pachytene chromosomes. Utilizing deletion stocks derived from irradiation experiments, marker loci were assigned to their respective positions in the pachytene chromosomes (Khush and Rick 1968). These same experiments led to integration of centromeres and other major cytological features into the genetic linkage map. It has been this combination of genetic and physical mapping that has made tomato unique among plants in terms of its development as a genetic organism (Rick 1975).

In the mid and late 1970's isozymes began to play a prominent role in linkage mapping in tomato and by 1980 a rough map, based on isozyme markers, had been constructed (Tanksley and Rick 1980). Because of their neutral phenotypes and codominance, isozymes could be utilized in experiments in which morphological markers would have been impractical. It was during this period, the first studies utilizing molecular markers for mapping quantitative trait loci (QTL) and 'gene tagging' were published (Rick and Fobes 1974; Tanksley et al. 1982).

By the mid 1980's RFLPs began to play a role in the genetic mapping of tomato. The first RFLP map was published in 1986 and this was followed by a number of articles reporting use of RFLPs in various aspects of breeding, genetics and evolutionary studies (Bernatzky and Tanksley 1986; Table 1). Because the RFLP linkage map of tomato was developed at an earlier time than most other organisms, it became the first species in which quantitative trait loci (QTLs) could be detected throughout an entire genome in a single segregating population (Paterson et al. 1988).

Table 1 Genes of known function or phenotype that have been mapped onto the molecular map of tomato/potato (morph = morphological marker)

Gene	Type	Product/phenotype	Chrm	Reference
6Pgdh-1	isozyme	6phosphogluconate dehydrogenase	4	Tanksley and Kuehn (1985), Bernatzky and Tanksley (1986)
6Pgdh-2	isozyme	6phosphogluconate dehydrogenase	12	Tanksley and Kuehn (1985)
6Pgdh-3	isozyme	6phosphogluconate dehydrogenase	5	Tanksley and Kuehn (1985), Bonierbale et al (1988)
a	morph	anthocyaninless	11	Rick (1980), Grandillo and Tanksley (unpub data)
ae	morph	entirely anthocyaninless	8	Rick (1980), Grandillo and Tanksley (unpub data)

Table 1. Continued.

Gene	Type	Product/phenotype	Chrm	Reference
af	morph	anthocyanin free	5	Rick (1980); Grandillo and Tanksley (unpub. data)
ag	morph	anthocyanin gainer	10	Rick (1980); Grandillo and Tanksley (unpub. data)
alb	morph	albescent	12	Rick (1980); Grandillo and Tanksley (unpub. data)
ACC1	RFLP	ACC synthase	8	Rottman et al. (1991)
ACC2	RFLP	ACC synthase	1	Rottman et al. (1991)
ACC3	RFLP	ACC synthase	2	Rottman et al. (1991)
ACC4	RFLP	ACC synthase	5	Rottman et al. (1991)
Aco-1	isozyme	aconitase	12	Tanksley and Rick (1980); Berntazky and Tanksley (1986)
Aco-2	isozyme	aconitase	7	Tanksley and Rick (1980); Berntazky and Tanksley (1986)
Adh-1	isozyme	alcohol dehydrogenase	4	Tanksley and Rick (1980); Berntazky and Tanksley (1986)
Adh-2	isozyme	alcohol dehydrogenase	6	Tanksley and Jones (1981); this report
Aps-1	isozyme	acid phosphatase	6	Tanksley and Rick (1980); this report
Aps-2	isozyme	acid phosphatase	8	Tanksley and Rick (1980); this report
B	morph	beta carotene	6	Rick (1980); Grandillo and Tanksley (unpub. data)
CAB1	RFLP	chlorophyll a/b binding polypeptide	2	Vallejos et al. (1986); Berntakzy and Tanksley (1986)
CAB2	RFLP	chlorophyll a/b binding polypeptide	8	Vallejos et al. (1986); Berntakzy and Tanksley (1986)
CAB3	RFLP	chlorophyll a/b binding polypeptide	3	Vallejos et al. (1986); Berntakzy and Tanksley (1986)
CAB4	RFLP	chlorophyll a/b binding polypeptide	7	Pichersky et al. (1987a); this report
CAB5	RFLP	chlorophyll a/b binding polypeptide	12	Pichersky et al. (1987a); Tanksley et al. (unpub. data)
CAB6	RFLP	chlorophyll a/b binding polypeptide	5	Pichersky et al. (1987b); this report
CAB7	RFLP	chlorophyll a/b binding polypeptide	10	Pichersky et al. (1988); this report
CAB8	RFLP	chlorophyll a/b binding polypeptide	10	Pichersky et al. (1989); this report
CAB11	RFLP	chlorophyll a/b binding polypeptide	6	Schwartz et al. (1991)
CAB12	RFLP	chlorophyll a/b binding polypeptide	3	Schwartz et al. (1991)
Cf-2	morph	resistance to <i>Cladosporium fulvum</i>	6	Jones et al. (1991)
Cf-9	morph	resistance to <i>Cladosporium fulvum</i>	1	Jones et al. (1991)
CHS1	RFLP	chalcone synthase	6	Drew and Goldberg (pers. comm.)

Table 1 Continued

Gene	Type	Product/phenotype	Chrm	Reference
CHS3	RFLP	chalcone synthase	5	Drew and Goldberg (pers comm)
CHS4	RFLP	chalcone synthase	9	Drew and Goldberg (pers comm)
E4	RFLP	ethylene inducible polypeptide	3	Lincoln et al (1987), this report
E8A	RFLP	ethylene inducible polypeptide	9	Lincoln et al (1987), this report
E8B	RFLP	ethylene inducible polypeptide	3	Lincoln et al (1987), this report
Est-1	isozyme	esterase	2	Tanksley and Rick (1980), Berntazky and Tanksley (1986)
Est-2	isozyme	esterase	9	Tanksley and Rick (1980), this report
Est-4	isozyme	esterase	12	Tanksley and Rick (1980), Berntazky and Tanksley (1986)
Est-5	isozyme	esterase	2	Tanksley and Rick (1980), Berntazky and Tanksley (1986)
Est-6	isozyme	esterase	2	Tanksley and Rick (1980), Berntazky and Tanksley (1986)
Est-7	isozyme	esterase	2	Tanksley and Rick (1980), Berntazky and Tanksley (1986)
Got-1	isozyme	glutamate oxaloacetate transaminase	4	Tanksley and Rick (1981)
Got-2	isozyme	glutamate oxaloacetate transaminase	7	Tanksley and Rick (1980), Berntazky and Tanksley (1986)
Got-3	isozyme	glutamate oxaloacetate transaminase	7	Tanksley and Rick (1980), Berntazky and Tanksley (1986)
h	morph	hairs absent	10	Rick (1980), Grandillo and Tanksley (unpub data)
hl	morph	hairless	11	Rick (1980), Grandillo and Tanksley (unpub data)
HMG2	RFLP	HMG CoA reductase	2	Narita and Gruissem, in prep
HMG3	RFLP	HMG CoA reductase	3	Narita and Gruissem, in prep
HTS1A	RFLP	heat shock transaction factor	2	Scharf et al (1990)
HTS1B	RFLP	heat shock transaction factor	2	Scharf et al (1990)
HTS1C	RFLP	heat shock transaction factor	10	Scharf et al (1990)
HTSB	RFLP	heat shock transaction factor	2	Scharf et al (1990)
HTS24	RFLP	heat shock transaction factor	8	Scharf et al (1990)
HTS30	RFLP	heat shock transaction factor	8	Scharf et al (1990)

Table 1 Continued

Gene	Type	Product/phenotype	Chrm	Reference
hy	morph	homogeneous yellow	10	Rick (1980), Grandillo and Tanksley (unpub data)
I2	morph	resistance to <i>Fusarium oxysporum</i> race 2	11	Safartti et al (1989)
I3	morph	resistance to <i>Fusarium oxysporum</i> race 3	7	Bournival et al (1989), Tanksley and Costello (1991)
Idh-1	isozyme	isocitrate dehydrogenase	1	Bernatzky and Tanksley (1986)
j	morph	jointless	11	Rick (1980), Wing and Tanksley (unpub data)
l-2	morph	lutescent-2	10	Rick (1980), Grandillo and Tanksley (unpub data)
Mdh-3	isozyme	malate dehydrogenase	7	Tanksley et al (unpub data)
Mi	morph	resistance to root knot nematodes	6	Messegeur et al (1991)
nor	morph	non-ripening	10	Rick (1980), Giovanonni and Tanksley (unpub data)
Nr	morph	never ripe	9	Rick (1980), Giovanonni and Tanksley (unpub data)
PGAL	RFLP	polygalaturonidase	10	Kinzer et al (1990), this report
Pgi-1	isozyme	phosphoglucosomerase (cytosolic)	12	Tanksley and Rick (1981), Bernatzky and Tanksley (1986)
Pgm-1	isozyme	phosphoglucosomutase (plastid)	3	Bernatzky and Tanksley (1986)
Pgm-2	isozyme	phosphoglucosomutase (cytosolic)	4	Tanksley and Rick (1981), Bernatzky and Tanksley (1986)
PPOIII	RFLP	polyphenol oxidase	8	Newman et al , submitted
PPOIVa	RFLP	polyphenol oxidase	12	Newman et al , submitted
PPOIVb	RFLP	polyphenol oxidase	4	Newman et al , submitted
Prx-1	isozyme	peroxidase	1	Tanksley and Rick (1980), Bernatzky and Tanksley (1986)
Prx-2	isozyme	peroxidase	2	Tanksley and Rick (1980), Bernatzky and Tanksley (1986)
Prx-3	isozyme	peroxidase	2	Tanksley and Rick (1980), Bernatzky and Tanksley (1986)
Prx-7	isozyme	peroxidase	3	Tanksley and Rick (1980), Bernatzky and Tanksley (1986)
PTC	RFLP	phytochrome	10	Lissemore et al (1987)
PTN	RFLP	patatin (tuber storage protein)	8	Bonierbale et al (1988), Ganai et al (1991)
Pto	morph	resistance to <i>Pseudomonas syringae</i>	5	Martin et al (1991)
rin	morph	ripening inhibitor	5	Giovanonni and Tanksley (unpub data)
R45s	RFLP	45s ribosomal RNA	2	Vallejos et al (1986), Bernatzky and Tanksley (1986)
R5s	RFLP	5s ribosomal RNA	1	Lapitan et al (1991)

Table 1 Continued

Gene	Type	Product/phenotype	Chrm	Reference
RBCS3	RFLP	ss ribulose bisphosphate carboxylase	2	Vallejos et al (1986), Berntakzy and Tanksley (1986)
RBCS2	RFLP	ss ribulose bisphosphate carboxylase	3	Vallejos et al (1986), Berntakzy and Tanksley (1986)
RBCS1	RFLP	ss ribulose bisphosphate carboxylase	2	Vallejos et al (1986), Berntakzy and Tanksley (1986)
Skdh-1	isozyme	shikimic acid dehydrogenase	1	Bernatzky and Tanksley (1986)
Sm	morph	resistance to <i>Stemphylum</i>	11	Behare et al (1991)
Sod-2	isozyme	superoxide dismutase	1	D Zamir et al (unpub data)
sp	morph	self-pruning	6	Paterson et al (1988)
spa	morph	sparsa	8	Rick (1980), Grandillo and Tanksley (unpub data)
tf	morph	trifoliolate	5	Rick (1980), Martin and Tanksley (unpub data)
Tm-1	morph	resistance to tobacco mosaic virus	2	Levesque et al (1990)
Tm-2a	morph	resistance to tobacco mosaic virus	9	Young et al (1988)
TOM25A	RFLP	ripening related	6	Kinzer et al (1990)
TOM25B	RFLP	ripening related	6	Kinzer et al (1990)
Tpi-2	isozyme	triose phosphate isomerase	4	Tanksley and Rick (1981), Bernatzky and Tanksley (1986)
Ve	morph	resistance to <i>Verticillium</i>	7	Juvick et al (1991)
Wx	RFLP	waxy	8	Gebhardt et al (1989)
y	morph	yellow flesh (potato)	3	Bonierbale et al (1988)

2. Current status of tomato molecular linkage map

Currently more than 1000 markers, covering a total of 1276 map units, have been mapped onto the tomato molecular linkage map (Tanksley et al. 1992; Fig. 1). Most of the markers correspond to random genomic clones or cDNA clones; however, a number of clones of known function have also been mapped (Table 2). Several morphological or isozyme markers from each chromosome have also been located on the molecular map which allows orientation of this map with respect to the classical linkage map (Tanksley et al. 1992; Fig. 1).

A novel feature of the tomato molecular map is that many of the markers on this map have also been mapped onto the potato molecular map. The result is that the maps for both species can be oriented with a high degree of precision

Table 2. Summary of applications of tomato molecular linkage map.

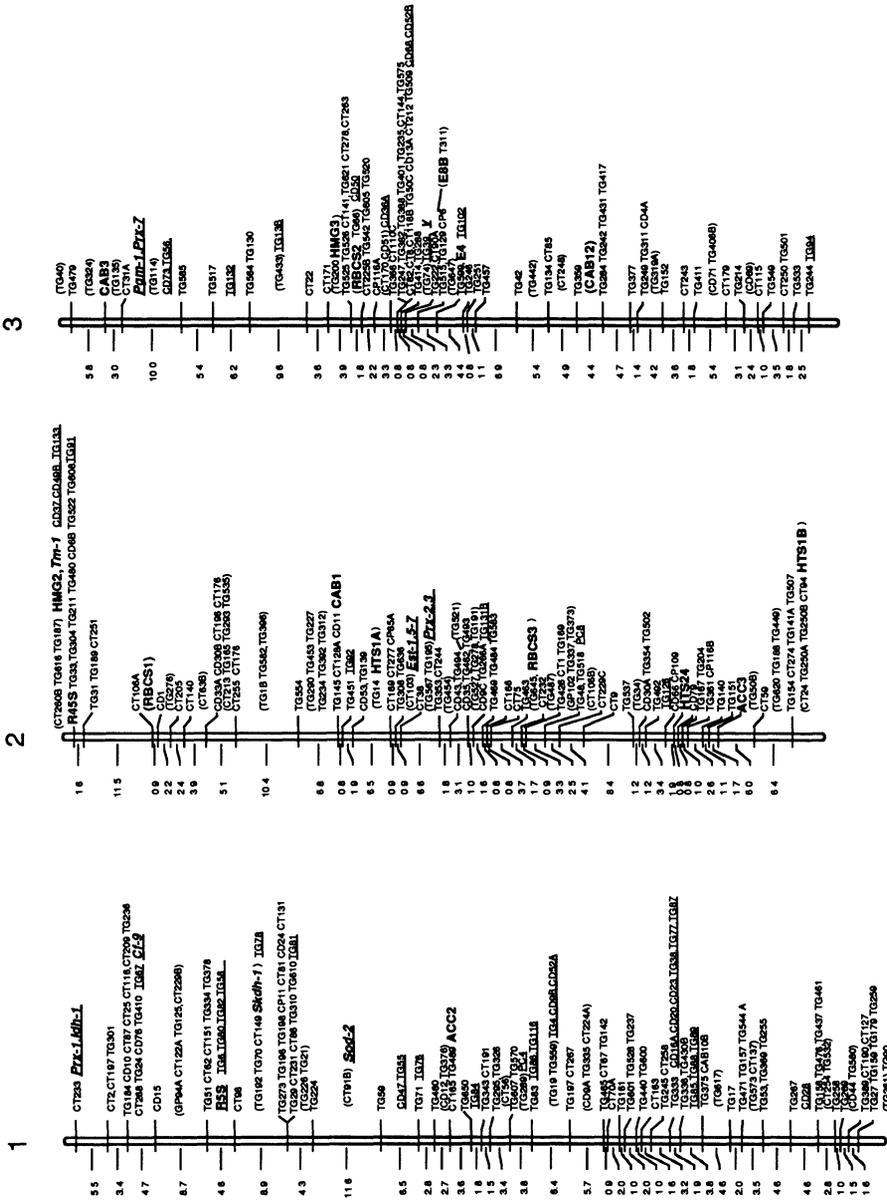
Quantitative genetics	Paterson et al. (1988) Paterson et al. (1989) Paterson et al. (1990)
Gene tagging ^a	Sarfatti et al. (1990) Messeguer et al. (1991) Martin et al. (1991) Tanksley and Costello (1991) Young et al. (1988) Giovanonni et al. (1991) Juvick et al. (1991) Behare et al. (1991) Jones et al. (1991) Van der Beek et al. (1991)
Evolutionary studies	Miller et al. (1990)
Genome organization/cytogenetics	DeVerna et al. (1990) Zamir and Tanksley (1988) Ganal et al. (1988) Lapitan et al. (1991) DeVicente and Tanksley (1992)
Comparative gene mapping	Tanksley et al. (1989, 1992) Bonierbale et al. (1988)
Plant breeding	Young and Tanksley (1989a,b)
Somatic hybrids	Bonnema et al. (1991) Wijbrandi et al. (1990) Bonnema and O'Connell (1992)
Physical mapping	Ganal et al. (1989, 1990)
YAC library characterization	Martin et al. (1992)
Methylation studies	Messeguer et al. (1991)
Transposon mapping	Osborne et al. (1991)

^a Establishment of tight linkage associations between major genes and one or more molecular markers.

(Bonierbale et al. 1988; Tanksley et al. 1992). Tomato and potato differ for 5 chromosomal arm translocations and the breakpoint of these ancestral locations can be pinpointed rather precisely on both maps (Tanksley et al. 1992). Because of the high degree of synteny between these two species, it is possible to utilize molecular probes from the potato map on tomato and vice versa. Besides the > 1000 markers mapped directly onto tomato, another approximately 300 markers have been mapped only in potato and their

→

Fig. 1. Molecular linkage map of the tomato genome. Loci by tick marks ordered with LOD > 3. Loci in bold correspond to known genes (see Table 1 for details). Loci following commas cosegregate. Markers enclosed in parentheses have been located to corresponding intervals with LOD < 3. Position of underlined loci approximated from placement on previously published maps. All other loci mapped directly on F₂ population of 67 plants from *L. esculentum* × *L. pennellii* (from Tanksley et al. 1992).



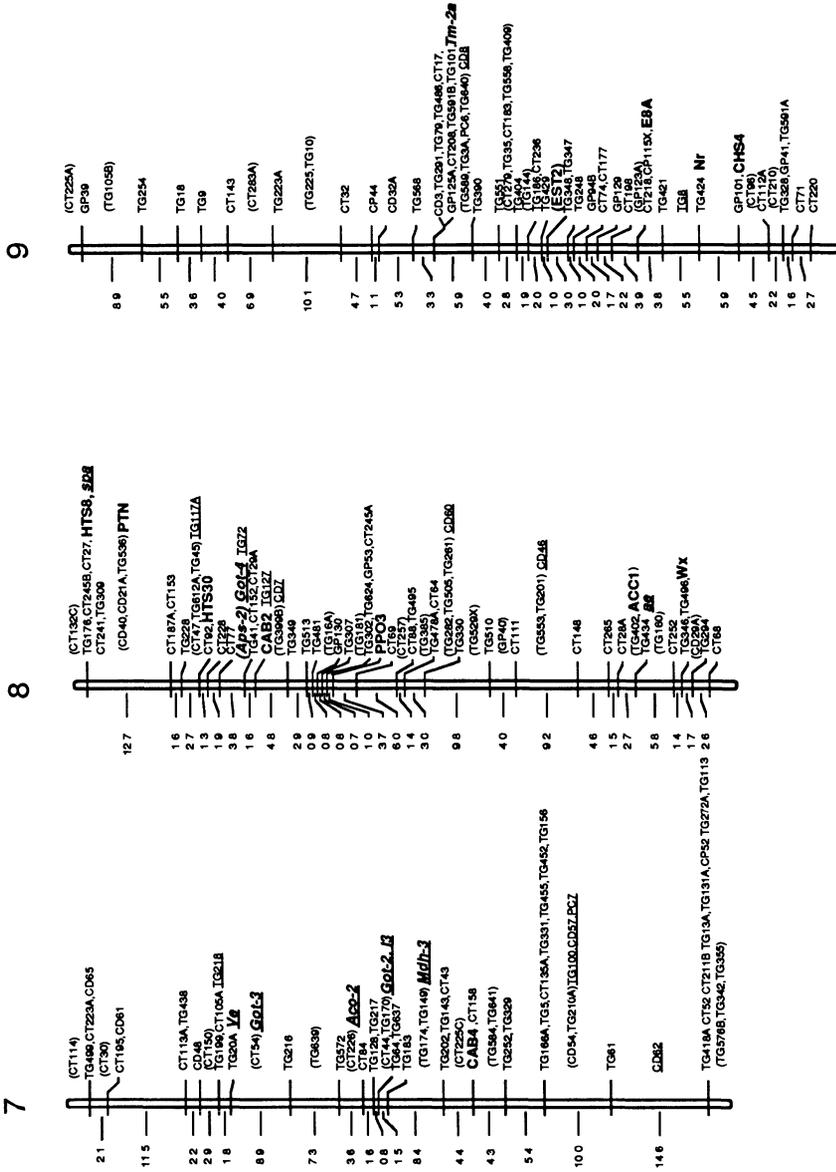


Fig. 1. Continued.

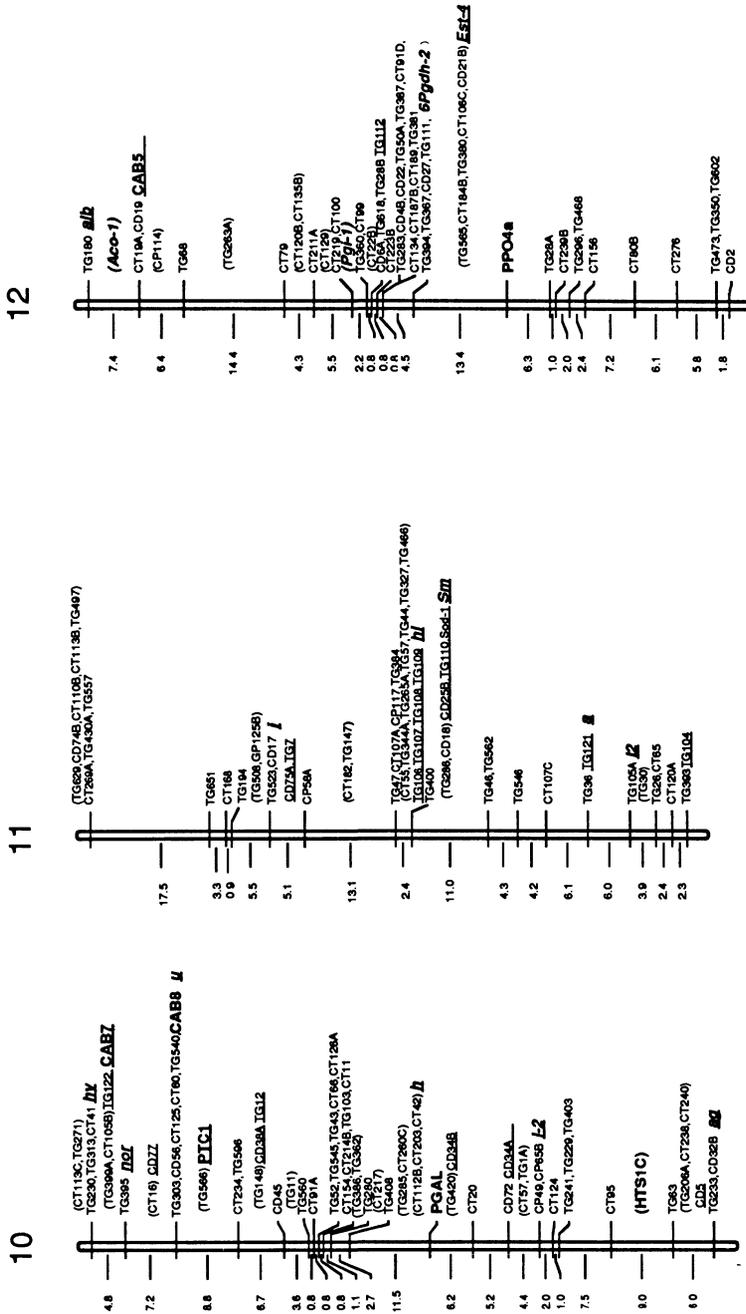


Fig. 1. Continued.

positions in tomato can be inferred with a high degree of precision (Gebhardt et al. 1989, 1991; pers. comm.). This brings the total number of markers available for tomato genetics to approximately 1300 which corresponds to (on average) 1 marker every 1 cM.

2.1. Localization of telomeres

Four of the tomato telomeres (chromosomes 5, 9, 11, 12) have recently been mapped onto the molecular linkage map. This was accomplished using pulsed field gel electrophoresis and a telomere-associated tandemly repeated DNA sequence (TGRI) (Ganal et al. 1988, 1992). The TGRI repeats at the ends of the chromosomes change rapidly in copy number, apparently due to unequal crossing over, resulting in polymorphism that can be visualized by Southern blotting of pulsed field gels (Broun et al. 1992). These polymorphisms are inherited in a Mendelian fashion and can be readily mapped relative to other markers. The fact that these telomeric polymorphisms map close (average distance 5 cM) to the most distal RFLP markers on the linkage map indicate that telomeres do not experience excessive rates of recombination and that the current tomato molecular linkage map covers the majority of the genome.

2.2. Heterogeneity of crossing over along the chromosomes

The distribution of map units within tomato chromosomes varies dramatically, depending on which part of the linkage group is being examined. This heterogeneity can be visualized by the clustering of larger numbers of markers in certain regions of the linkage groups (Fig. 1). Normally, this clustering occurs at a position in the linkage group that corresponds to the approximate position of the centromeres (Tanksley et al. 1992). Histograms of marker density per map unit for chromosomes 3 and 12 clearly demonstrate this point (Fig. 2). That the clustering of markers around centromeres is due to reduced levels of crossing over and not other factors is supported by physical mapping studies using pulsed field gel electrophoresis (Ganal et al. 1989). Physical mapping indicates a suppression of recombination around centromeres of at least 7 fold compared with the rest of the chromosome. Marker density heterogeneity suggests a suppression in recombination around centromeres of approximately 10 fold and is in remarkable agreement with the physical mapping data (Tanksley et al. 1992).

3. Future prospects

In recent years there has been an effort to construct linkage maps in which the markers are spaced at very close intervals throughout the genome. Such high density maps can serve a number of purposes in basic and applied research including chromosome walking, marker-based selection of desirable genes in

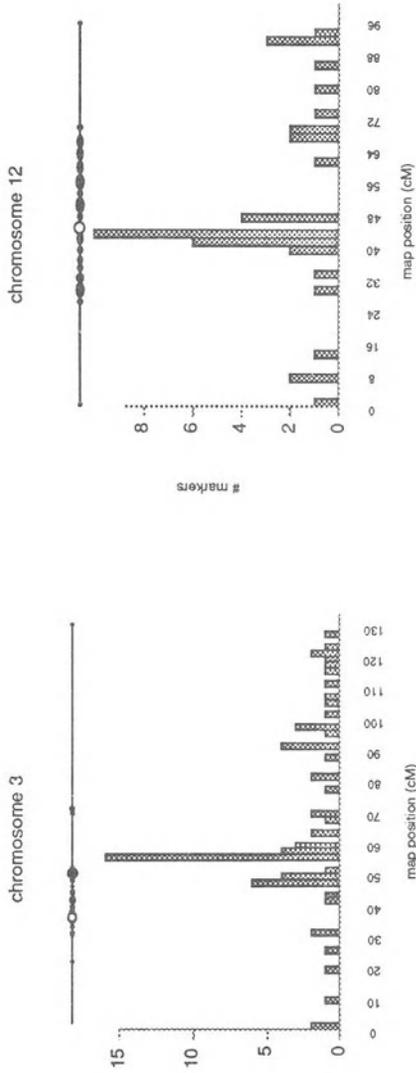


Fig. 2. Histograms showing density of markers (per 2 cM intervals) along chromosomes 3 and 12 (from Tanksley et al. 1992).

breeding programs and quantitative genetics studies. Currently the tomato molecular map is being used for all of these purposes. One of the potentially most powerful applications of the tomato map is in chromosome walking or map-based cloning. Because the markers on the map are at very close intervals and since the genome is relatively small among crop plants, tomato represents an ideal crop species for map-based cloning. Other features that make tomato especially amenable to map-based cloning is the relative ease with which it can be transformed with exogenous DNA and the availability of a yeast artificial chromosome library (Martin et al. 1992). It is anticipated that in the future the molecular linkage map will play a key role in cloning genes for a variety of economically valuable and/or biologically interesting characters including disease resistance, growth habit, fruit ripening behavior and even yield.

4. References

- Arumuganathan, K and Earle, E D (1991) Nuclear DNA content of some important plant species *Plant Mol Biol Rep* 9 208–218
- Behare, J , Laterrot, H , Sarfatti, M and Zamir, D (1991) RFLP mapping of the *Stemphylium* resistance gene in tomato *Mol Plant-Microbe Interactions* 4 489–492
- Bernatzky, R and Tanksley, S D (1986) Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences *Genetics* 112 887–898
- Bonierbale, M , Plaisted, R L and Tanksley, S D (1988) RFLP maps of potato and tomato based on a common set of clones reveal modes of chromosomal evolution *Genetics* 120 1095–1103
- Bonnema, A B and O'Connell, M A (1992) Molecular analysis of the nuclear organellar genotype of somatic hybrid plants between tomato (*Lycopersicon esculentum*) and *Lycopersicon chilense* *Plant Cell Rep* 10 629–632
- Bonnema, A B , Melzer, J M and O'Connell, M A (1991) Tomato cybrids with mitochondrial DNA from *Lycopersicon pennellii* *Theor Appl Genet* 81 339–348
- Bournival, B L , Vallejos, C E and Scott, J W (1989) An isozyme marker for resistance to race 3 of *Fusarium oxysporum* f sp *lycopersici* in tomato *Theor Appl Genet* 78 489–494
- Broun, P , Ganai, M G and Tanksley, S D (1992) Telomeric arrays display high levels of heritable polymorphisms among closely related plant varieties *Proc Natl Acad Sci U S A* 89 1354–1357
- Butler, L (1951) New linkage groups in the tomato *J Hered* 42 100–104
- Clayberg, C D , Butler, L , Kerr, E A , Rick, C M and Robinson, R W (1966) Third list of known genes in the tomato *J Hered* 57 189–197
- DeVerna, J W , Rick, C M , Chetelat, R T , Lamni, B J and Alpert, K B (1990) Sexual hybridization of *Lycopersicon esculentum* and *Solanum rickii* by means of a sesquidiploid bridging hybrid *Proc Natl Acad Sci U S A* 87 9486–9490
- DeVicente, M C and Tanksley, S D (1991) Genome-wide reduction in recombination in progeny derived from male versus female gametes in an interspecific cross of tomato *Theor Appl Genet* 83 173–178
- Ganai, M W , Lapitan, N L V and Tanksley, S D (1988) A molecular and cytogenetic survey of major repeated DNA sequences in tomato (*Lycopersicon esculentum*) *Mol Gen Genet* 213 262–268
- Ganai, M W , Young, N D and Tanksley, S D (1989) Pulsed field gel electrophoresis and physical mapping of large DNA fragments in the Tm-2a region of chromosome 9 in tomato *Mol Gen Genet* 215 395–400
- Ganai, M W , Bonierbale, M W , Roeder, M S , Park, W D and Tanksley, S D (1990) Genetic and physical mapping of the patatin genes in potato and tomato *Mol Gen Genet* 225 501–509

- Ganal, M W , Lapitan, N L V and Tanksley, S D (1991) Macrostructure of the tomato telomeres *Plant Cell* 3 87-94
- Ganal, M W , Broun, P and Tanksley S D (1992) Genetic mapping of tandemly repeated telomeric DNA sequences in tomato (*Lycopersicon esculentum*) *Genomics* 14(2) 444-448
- Gebhardt, C , Ritter, E , Debener, T , Schachtsschabel, U , Walkemeier, B , Uhrig, H and Salamini, F (1989) RFLP analysis and linkage mapping in *Solanum tuberosum* *Theor Appl Genet* 78 65-75
- Gebhardt, C , Ritter, E , Barone, A , Debener, T , Walkemeier, B , Schachtschabel, U , Kaufman, H , Thompson, R D , Bonierbale, M W , Ganal, M W , Tanksley, S D and Salamini, F (1991) RFLP maps of potato and their alignment with the homoeologous tomato genome *Theor Appl Genet* 83 49-57
- Giovannoni, J , Wing, R and Tanksley, S D (1992) Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations *Nucl Acids Res* 19 6553-6558
- Jones, D A , Dickinson, M J and Jones, J D G (1991) Genetic locations of the Cf-2, Cf-4, Cf-5 and Cf-9 genes for resistance to *Cladosporium fulvum* *Rep Tomato Genet Coop* 41 22
- Juvick, J A , Bolkan, H and Tanksley, S D (1991) The Ve gene for race 1 *Verticillium* resistance is on chromosome 7 *Rep Tomato Genet Coop* 41 23-24
- Khush, G S and Rick, C M (1968) Cytogenetic analysis of the tomato genome by means of induced deficiencies *Chromosoma* 23 452-484
- Kinzer, S M , Schwager, S J and Mutschler, M A (1990) Mapping of ripening-related or -specific cDNA clones of tomato (*Lycopersicon esculentum*) *Theor Appl Genet* 79 489-496
- Lander, E S , Green, P , Abrahamson, J , Barow, A , Daly, M J , Lincoln, S E and Newburg, L (1987) MAPMAKER An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations *Genomics* 1 174-181
- Lapitan, N L V , Ganal, M W and Tanksley, S D (1989) Somatic chromosome karyotype of tomato based on in situ hybridization of the TGRI satellite repeat *Genome* 32 992-998
- Lapitan, N L V , Ganal, M W and Tanksley, S D (1991) Organization of the 5S ribosomal RNA genes in the genome of tomato *Genome* 34 509-514
- Levesque, H , Vedel, F , Mathieu, C and de Courcel, A G L (1990) Identification of a short rDNA spacer sequence highly specific to a tomato line containing the Tm-1 gene introgressed from *Lycopersicon hirsutum* *Theor Appl Genet* 80 602-608
- Lincoln, J E , Cordes, S , Read, E and Fischer, R L (1987) Regulation of gene expression by ethylene during *Lycopersicon esculentum* (tomato) fruit development *Proc Natl Acad Sci U S A* 84 2793-2797
- Lissemore, J L , Colbert, J T and Quail, P H (1987) Cloning of cDNA for phytochrome from etiolated *Cucurbita* and coordinate photoregulation of the abundance of two distinct phytochrome transcripts *Plant Mol Biol* 8 485-496
- Martin, G B , Williams, J G K and Tanksley, S D (1991) Rapid identification of markers near a *Pseudomonas* resistance gene in tomato using random primers and near-isogenic lines *Proc Natl Acad Sci U S A* 88 2336-2340
- Martin, G B , Ganal, M W and Tanksley, S D (1992) Construction of a yeast artificial chromosome library of tomato and identification of cloned segments linked to two disease resistance loci *Mol Gen Genet* 233 25-32
- Messeguer, R , Ganal, M W , deVicente, M C , Young, N D , Bolkan, H D and Tanksley, S D (1991) High resolution RFLP map around the root-knot nematode resistance gene (Mi) in tomato *Theor Appl Genet* 82 529-536
- Messeguer, T , Ganal, M W , Steffens, J C and Tanksley, S D (1991) Characterization of the level, target sites and inheritance of cytosine methylation in tomato nuclear DNA *Plant Mol Biol* 16 753-770
- Miller, J C and Tanksley, S D (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon* *Theor Appl Genet* 80 437-448

- O'Connell, M E and Hanson, M K (1986) Regeneration of somatic hybrid plants formed between *Lycopersicon esculentum* and *Solanum rickii* *Theor Appl Genet* 72 59-65
- Osborne, B I , Corr, C , Prince, J , Hehl, R , Tanksley, S D , McCormick, S and Baker, B (1990) Ac transposition from a T-DNA can generate linked and unlinked clusters of insertions in the tomato genome *Genetics* 129 833-844
- Paterson, A H , Lander, E S , Hewitt, J D , Peterson, S , Lincoln, S E and Tanksley, S D (1988) Resolution of quantitative traits into Mendelian factors, using a complete linkage map of restriction fragment length polymorphism *Nature* 335 721-726
- Paterson, A H , DeVerna, J W , Lanini, B and Tanksley, S D (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato *Genetics* 124 735-742
- Paterson, A H , Damon, S , Hewitt, J D , Zamir, D , Rabinowitch, H D , Lincoln, S E , Lander, E S and Tanksley, S D (1991) Mendelian factors underlying quantitative traits in tomato comparison across species, generations, and environments *Genetics* 127 181-197
- Pichersky, E , Hoffman, N E , Malik, V S , Bernatzky, B , Tanksley, S D , Szabo, L and Cashmore, A R (1987) The tomato Cab-4 and Cab-5 genes encode a second type of CAB polypeptides localized in Photosystem II *Plant Mol Biol* 9 109-120
- Pichersky, E , Tanksley, S D , Piechulla, B , Stayton, M M and Dunsmuir, P (1988) Nucleotide sequence and chromosomal location of *Cab-7*, the tomato gene encoding the Type II chlorophyll a/b-binding polypeptide of Photosystem I *Plant Mol Biol* 11 69-71
- Pichersky, E , Brock, T G , Nguyen, D , Hoffman, N E , Piechulla, B , Tanksley, S D and Green, B R (1989) A new member of the CAB gene family structure, expression and chromosomal location of *Cab-8*, the tomato gene encoding the Type III chlorophyll a/b binding polypeptide of photosystem I *Plant Mol Biol* 12 257-270
- Rick, C M (1975) The tomato In R C King (ed), *Handbook of Genetics*, Vol 2, pp 247-280 Plenum Press, New York
- Rick, C M (1980) Tomato linkage survey *Rept Tomato Genet Coop* 30 2-17
- Rick, C M and Fobe, J F (1974) Association of an allozyme with nematode resistance *Rep Tomato Genet Coop* 24 25
- Rick, C M and Yoder, J (1988) Classical and molecular genetics of tomato highlights and perspectives *Ann Rev Genet* 22 281-300
- Rottman, W H , Peter, G F , Oeller, P W , Keller, J A , Shen, N F , Nagy, B P , Taylor, L P , Campbell, A D and Theologis, A (1991) 1-aminocyclopropane-1 carboxylate synthase in tomato is encoded by a multigene family whose transcription is induced during fruit and floral senescence *J Mol Biol* 222 937-961
- Sarfatti, M , Katan, J , Fluhr, R and Zamir, D (1989) An RFLP marker in tomato linked to the *Fusarium oxysporum* resistance gene I2 *Theor Appl Genet* 78 755-759
- Scharf, K , Rose, S , Zott, W , Schoff, F and Nover, L (1990) Three tomato genes code for heat stress transcription factors with a region of remarkable homology to the DNA-binding domain of the yeast HSF *EMBO J* 9 4495-4501
- Schwartz, E , Shen, D , Aebersold, R , McGrath, J M , Pichersky, E and Green, B R (1991) Nucleotide sequence and chromosomal location of Cab11 and Cab12, the genes for the fourth polypeptide of the photosystem I light-harvesting antenna (LHCI) *FEBS Lett* 280 229-234
- Tanksley, S D and Costello, W (1991) The size of the *L pennellii* chromosome 7 segment containing the I-3 gene in tomato breeding lines measured by RFLP probing *Rept Tomato Genet Coop* 41 60
- Tanksley, S D and Jones, R A (1981) Effect of O2 stress on tomato ADH's description of a second ADH coding gene *Biochem Genet* 19 397
- Tanksley, S D and Kuehn, G (1985) Genetics, subcellular localization and molecular characterization of 6-phosphogluconate dehydrogenase isozymes in tomato *Biochem Genet* 23 442-454
- Tanksley, S D and Rick, C M (1980) Isozymic gene linkage map of the tomato Applications in genetics and breeding *Theor Appl Genet* 57 161-170

- Tanksley, S.D., Medina-Filho, H. and Rick, C.M. (1982) Use of naturally occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. *Heredity* 49: 11–25.
- Tanksley, S.D., Young, N.D., Paterson, A.H. and Bonierbale, M.W. (1989) RFLP mapping in plant breeding: new tools for an old science. *Bio/Technology* 7: 257–264.
- Tanksley, S.D., Ganal, M.W., Prince, J.P., deVicente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., Messesguer, R., Miller, J.C., Miller, L., Paterson, A.H., Pineda, O., Roder, M., Wing, R.A., Wu, W. and Young, N.D. (1992) High density molecular linkage maps of the tomato and potato genomes: biological inferences and practical applications. *Genetics* 132: 1141–1160.
- Vallejos, C.E., Tanksley, S.D. and Bernatzky, R. (1986) Localization in the tomato genome of DNA restriction fragments containing sequences coding for rDNA (45s), ribulose biphosphate carboxylase and chlorophyll a/b binding protein. *Genetics* 112: 93–105.
- Van der Beek, H., Verkerk R. and Lindhout, P. (1991) Mapping gene Cf-9 (resistance to *Cladosporium fulvum*) on chromosome 1 of tomato. *Rep. Tomato Genet. Coop.* 41: 15.
- Wijbrandi, J., Zabel, P. and Koornneef, M. (1990) Restriction fragment length polymorphism analysis of somatic hybrids between *Lycopersicon esculentum* and irradiated *L. peruvianum*: evidence for limited donor genome elimination and extensive chromosome rearrangements. *Mol. Gen. Genet.* 222: 270–277.
- Young, N.D. and Tanksley, S.D. (1989a) RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. *Theor. Appl. Genet.* 77: 353–359.
- Young, N.D. and Tanksley, S.D. (1989b) Restriction fragment length polymorphism maps and the concept of graphical genotypes. *Theor. Appl. Genet.* 77: 95–101.
- Young, N.D., Zamir, D., Ganal, MW. and Tanksley, S.D. (1988) Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the *Tm-2a* gene in tomato. *Genetics* 120: 579–585.
- Zamir, D. and Tanksley, S.D. (1988) Tomato genome is comprised largely of fast-evolving low copy-number sequences. *Mol. Gen. Genet.* 213: 254–261.

21. RFLP maps of bread wheat

GARY E. HART

Soil and Crop Sciences Department, Texas A&M University, College Station, TX 77843, U.S.A.

Contents

1. Introduction	327	6. International Triticeae Mapping Initiative	354
2. RFLP nomenclature	328	7. Concluding remarks	354
3. RFLP loci	329	8. References	355
4. RFLP maps	345		
5. RFLP maps of wheat relatives	353		

1. Introduction

Triticum aestivum (bread wheat, $2n = 6x = 42$, genomes A, B and D) is poorly genetically mapped in comparison to most of the other important crop plant species. The characteristics that are the most significant causes of this are polyploidy, a large genome, and autogamy. Due to polyploidy, most genes of *T. aestivum* are triplicated or at least duplicated. Consequently, the potential morphological and physiological effects of many recessive or inactive alleles are not expressed because they are masked by the effects of dominant or active alleles at paralogous loci. The large number of chromosomes in wheat do not pose a significant hindrance to genetic mapping because the numerous and diverse aneuploid stocks that are available allow the chromosome and chromosome-arm locations of genes and RFLPs to be readily determined. However, the large amount of DNA in the wheat genome (about 18 pg per haploid nucleus) makes the techniques for study of DNA polymorphisms more difficult to perform than with most crop plant species. Autogamy is a major cause of the difficulty experienced in genetically mapping wheat; the low level of DNA restriction-fragment-length polymorphism and allozyme variation that exists among wheat varieties is the most serious impediment to mapping the species.

In spite of the aforementioned difficulties, progress is being made in mapping wheat. Methods have been devised that overcome the problems associated with studying RFLPs in a genome containing a large amount of DNA and, in the

recent past, wide crosses have been used to generate genetic variability in sufficient quantity to allow the construction of partial molecular-marker maps of many wheat chromosomes. While no single chromosome of wheat can yet be properly considered to be well-mapped, a large number of DNA fragment-clone combinations have been identified and the rate of genetic mapping of the species is accelerating. Consequently, it is likely that low resolution RFLP maps of all wheat chromosomes will be produced in the near future.

This chapter reviews the current status of wheat RFLP mapping. A table listing the chromosome and/or chromosome-arm locations of wheat RFLP loci and the DNA clones used to detect them is presented. The published RFLP linkage maps of wheat are presented and described and RFLP linkage studies of *T. tauschii*, the D-genome progenitor of *T. aestivum*, are briefly summarized. Also, the genetic nomenclature that has been adopted for RFLP loci and alleles of wheat and related species is described.

It should be noted that a complete catalogue of wheat genetic markers is published every five years and a supplement to the catalogue is published annually. The catalogue lists all known wheat genes and alleles, a prototype strain for each allele, the chromosomal locations of all known genetic markers (including RFLPs), genetic maps of chromosomes, and other information. The catalogue is published in the Proceedings of the International Wheat Genetics Symposium; for the most recent edition, see McIntosh, Hart and Gale (1993). The catalogue is also available on a computer disk. The annual supplements are prepared by R.A. McIntosh, M.D. Gale and the author and are published in Cereal Research Communications, the Wheat Information Service, and the Wheat Newsletter.

2. RFLP nomenclature

Nomenclature guidelines for RFLP loci and alleles of wheat and related species are presented in Hart and Gale (1988). The guidelines specify that the basic symbol for all loci that are detected by the study of RFLPs, with the exception of known protein and RNA loci, should be 'X'. For loci that are detected with anonymous probes, the 'X' is followed by a laboratory designator (consisting of three lower-case letters), a probe number (consisting of Arabic numerals assigned sequentially by the laboratory involved), a hyphen (-), and the symbol for the chromosome in which the locus is located. All characters in the locus symbol are italicized. RFLP loci located in different chromosomes that are detected with the same probe should be assigned the same symbol except for the chromosome designation. By way of example, *Xpsr115-5B* designates a RFLP locus located in chromosome 5B that is detected with probe 115 of the Institute of Plant Science Research (IPSR) and *Xpsr115-5D* and *Xpsr115-4A* designate loci in 5D and 4A, respectively, that are detected with the same probe. Two or more RFLP loci located in one chromosome that are detected with the same probe should be assigned the same symbol except for the addition of different

Arabic numerals in parentheses immediately after the chromosome designation. For example, *Xpsr120-5R(1)* and *Xpsr120-5R(2)* designate two of the several RFLP loci in chromosome 5R of *Secale cereale* that are detected with IPSR probe 120.

RFLP loci that are detected with 'known-function' probes and for which neither a product has been identified nor allelism with a known functional locus has been demonstrated are designated in the same manner as loci that are detected with anonymous probes except that an abbreviation of the name of the function is substituted for the laboratory designator and probe number. The first letter of the abbreviated name is capitalized and the other letter(s) are lower-case. For example, *XNra-7A* designates a RFLP locus in chromosome 7A that is identified with a nitrate reductase probe and for which no protein product has been demonstrated and *X α -Amy-7B* designates a RFLP locus in 7B that is identified with an α -amylase probe and for which allelism with the known functional α -amylase locus in 7B (*α -Amy-B2*) has not been demonstrated.

A locus detected by the study of RFLPs that is shown to be a known protein or RNA locus or whose function is subsequently determined should be designated with the appropriate protein or RNA locus symbol. Nomenclature guidelines for protein and RNA loci of wheat and related species are contained in Hart and Gale (1988).

Alleles at RFLP loci should be designated with the name of the restriction enzyme followed by a lower-case letter, with the allele present in *T. aestivum* cv. Chinese Spring designated 'a'. For example, *Xpsr117-7A-EcoRIa* denotes the *Xpsr117-7A* allele detected in Chinese Spring with *EcoRI*.

On genetic maps and in other contexts, symbols for RFLP loci and alleles may often be abbreviated. For example, on a map of 7A, *Xpsr117* may be used in place of *Xpsr117-7A* and, furthermore, if all of the anonymous probes associated with the map are IPSR probes, *X117* may be used.

Note added in proof: Revision of the nomenclature guidelines for DNA markers in wheat and related species was authorized at the 1993 International Wheat Genetics Symposium. The revisions will be published in the 1994 supplement to the Catalogue of Gene Symbols for Wheat. The principal change is that symbols for *all* loci of unknown function should consist of a 'X', a laboratory designator, a number, a hyphen, and the symbol for the chromosome in which the locus is located (as described above) but with the added provision that symbols for loci detected with 'known-function' DNA clones or with primers that amplify genes may include a gene symbol in parentheses immediately before the hyphen.

3. RFLP loci

Table 1 presents a summary of the *T. aestivum* RFLP loci that have been reported in publications to date. The chromosome-arm location or, if the arm location is not known, the chromosome location of each locus is shown. Also

listed are the clones used to detect the loci and the chromosomal locations of loci in other chromosomes that have been detected with the same clones.

With the exception of the *glk* loci reported by Liu and Tsunewaki (1991), the chromosomal locations of the loci listed in Table 1 were determined almost exclusively by using aneuploid lines, principally compensating nullisomic-tetrasomic lines, ditelosomic lines, and disomic wheat-alien chromosome addition lines. Some of the *glk* loci were assigned to chromosomes using aneuploids but the chromosomal assignments of most of the loci were based on linkage analyses.

The vast majority of the DNA clones that have been used to detect wheat RFLP loci are either anonymous cDNA clones or anonymous genomic DNA (gDNA) clones. Most of these clones were obtained from *T. aestivum*, a large number from *T. tauschii* and some from barley, oats and other species. A small number of known-function probes have been used to identify RFLP loci and in a few cases the loci identified with these probes have been shown to be functional loci that had been identified in earlier studies of proteins (e.g., the *Glu-1* genes located in the group 1L chromosome arms), isozymes (e.g., the β -*Amy-1* loci in 4BL, 4DL and 5AL) and nucleolar organizer regions (e.g., *Nor-B1* in 1BS). Most cDNA clones hybridize to fragments located in each of the three members of one or sometimes two homoeologous chromosome-arm groups; very few clones hybridize to fragments from only one or two chromosomes in a group or to chromosomes in more than two groups (Sharp et al. 1989; Chao et al. 1989a,b; Devos et al. 1992a). Low-copy-number anonymous gDNA clones hybridize less frequently than cDNA clones to sequences located in each of the three members of a homoeologous group. Liu and Tsunewaki (1991) found that only 32% of the 72 *Pst*I genomic clones that they tested did so and that 49% of the clones detected loci in the chromosomes of two or more groups.

Ten of the 15 homoeologous group 3 gDNA probes studied by Devos et al. (1992a) hybridized to one or more fragments in each of the three arms of one of the group 3 arm groups and the remainder hybridized either to fragments located in non-homoeologous chromosomes or to one or more fragments located in one chromosome only. Anderson et al. (1992) determined the chromosome-arm locations of over 800 DNA fragments using 210 barley and oat cDNA clones and wheat gDNA clones (the number of cDNA as versus gDNA clones studied was not reported). Seventy-seven percent of the clones hybridized to fragments from one chromosome-arm group only and almost all of these clones hybridized to at least one fragment in each of the three arms. This high frequency is undoubtedly due in good part to the fact that in this study, preference was given in choosing clones and the restriction enzymes to be used with them to clone-enzyme combinations that yielded only three fragments of approximately equal hybridization intensity.

It must be emphasized that Table 1 does not list all of the known wheat RFLP loci. It includes only 22 *cnl* loci detected with seven probes; the chromosomal locations of the many loci detected by the other 203 probes studied by Anderson et al. (1992) have not been published. Also, Liu and Tsunewaki (1991)

Table 1 RFLP loci of *Triticum aestivum* (listed by the chromosome arm group in which they are located or, if the arm location is unknown by chromosome group), synonyms for locus designations, clones, and locations of other RFLP loci detected with the clones

Locus [synonym(s)]	DNA clone(s)	Locations of other loci detected with clones
Group 1S^a		
<i>XGh-A,B,D1</i> ^b (36) ^c	pTag544	
<i>Gh-A,B,D1-1</i> (62, 63)	pTag 1436 (4)	
<i>XIca1-1A,B,D</i> (62)	pcI-1-4 (65)	
<i>Nor-B1</i> (2, 23, 48, 60)	pTa71 and derivates of pTa250 (30)	
<i>Xpsr161-1A,B,D</i> (57)	PSR 161	
<i>Xpsr168-1A,B,D</i> (62, 63)	PSR 168	
<i>Xpsr381-1A,B,D</i> (62)	PSR 381	
<i>Xpsr393-1A,B,D</i> (62, 63)	PSR 393	
<i>Xpsr596-1A,B,D</i> (62)	PSR 596	
<i>Xpsr688-1A,B,D</i> (62)	PSR 688	
<i>Xpsr908-1B</i> (21)	PSR 908	(2A,D,6B)
<i>5S-Rrna-B1</i> (3, 24)	pTa794 (31)	
[<i>5S-Dna-B1</i> (24)]		
<i>5S-Rrna-D1</i> (24)	pTa794 (31)	
[<i>5S-Dna-D1</i> (24)]		
Group 1L		
<i>XAdh-1A,B,D</i> (50)	p3NTR [50]	
[<i>Adh-A,B,D2</i> (50), <i>Xadh3'</i> (32)]		
<i>Xcn1CDO1312-1B</i> (1)	CDO1312	(4B,D,5A)
<i>XEm-1A,B,D</i> (27)	p10-15	
<i>Glu-A,B,D1-1</i> (37,61)	pTag1290 (61)	
<i>Glu-A,B,D1-2</i> (37,61)	pTag1290 (61)	
<i>XLec-1A,B,D</i> (62, 63)	PNVR1 (51)	
<i>XPgk1-1A,B,D</i> (12)	P7 (46)	
<i>XPpdk-1A,B,D</i> (12)	PPDK4 (64)	
<i>Xpsr121-1A,B,D</i> (28)	PSR 121	(7A,B,D)
<i>Xpsr158-1A,B,D</i> (62, 63)	PSR 158	
<i>Xpsr159-1A,B,D</i> (62, 63)	PSR 159	
<i>Xpsr162-1A,B,D</i> (57)	PSR 162	
<i>Xpsr325-1A,B,D</i> (63)	PSR 325	
<i>Xpsr330-1A,B,D</i> (62, 63)	PSR 330	
<i>Xpsr343-1A,B,D</i> (62)	PSR 343	
<i>Xpsr361-1A,B,D</i> (62)	PSR 361	
<i>Xpsr385-1A,B,D</i> (62)	PSR 385	
<i>Xpsr391-1A,B,D</i> (62, 63)	PSR 391	
<i>Xpsr549-1A</i> (19, 21)	PSR 549	(2B,3A)
<i>Xpsr586-1A,B,D</i> (62)	PSR 586	
<i>Xpsr626-1A,B,D</i> (62)	PSR 626	
<i>Xpsr653-1A,B,D</i> (62)	PSR 653	

Table 1. Continued.

Locus [synonym(s)]	DNA clone(s)	Locations of other loci detected with clones
Group 1		
<i>Xgk190-1B</i> (45)	pTag90	
<i>Xgk194-1D</i> (45)	pTag94	(2,3D,6A)
[<i>Xgk194d</i> (45)]		
<i>Xgk136-1B</i> (45)	pTag136	
<i>Xgk163-1B</i> (45)	pTag163	
<i>Xgk427-1B</i> (45)	pTag427	
<i>Xgk483-1B</i> (45)	pTag483	
<i>Xgk520-1B</i> (45)	pTag520	(2,5A,3,6B)
[<i>Xgk520a</i> (45)]		
<i>Xgk549-1B</i> (45)	pTag549	(7B)
[<i>Xgk549b</i> (45)]		
<i>Xgk558-1D</i> (45)	pTag558	
<i>Xgk595-1B</i> (45)	pTag595	(3A)
[<i>Xgk595b</i> (45)]		
<i>Xgk652-1D</i> (45)	pTag652	(3B)
[<i>Xgk652a</i> (45)]		
<i>Xgk710-1A</i> (45)	pTag710	
<i>Xgk732-1A</i> (45)	pTag732	
<i>Xgk764-1B</i> (45)	pTag764	
<i>XksuD3-1D</i> (32)	pTtksuD3	
<i>XksuD16-1D</i> (32)	pTtksuD16	(<i>Ae. squarrosa</i> 5D)
<i>XksuD40-1D</i> (32)	pTtksuD40	
<i>XksuD49-1D</i> (32)	pTtksuD49	
<i>Xksu123-1D</i> (32)	pTtksu123	
<i>XksuM88-1D</i> (32)	pTtksuM88	
Group 2S		
<i>Xβ-Amy-2A,B,D</i> (58)	pcbC51 (41)	
[<i>PSR1-2A,B,D</i> (58)]		
<i>Xpsr-2A,B,D</i> (21)	POX375 (54)	
<i>Xpsr100-2A,B,D</i> (21)	PSR100	(5A,B,D)
<i>Xpsr107-2A,B,D</i> (21)	PSR107	
<i>Xpsr108-2A,B,D</i> (21, 56)	PSR108	(7A,B,D)
<i>Xpsr109-2A,B,D</i> (21, 56)	PSR109	(5A,B,D)
<i>Xpsr126-2A,B,D</i> (21)	PSR126	
<i>Xpsr130-2A,B,D</i> (21, 56)	PSR130	
<i>Xpsr131-2A,B,D</i> (21, 56)	PSR131	
<i>Xpsr135-2A,B,D</i> (21, 57)	PSR135	
<i>Xpsr137-2A,B,D</i> (21)	PSR137	
<i>Xpsr143-2A</i> (21)	PSR143	
<i>Xpsr146-2A,B,D</i> (21)	PSR146	
<i>Xpsr150-2A,B,D</i> (21)	PSR150	(5A,B,D,7A,B,D)
<i>Xpsr379-2A,B,D</i> (21)	PSR379	
<i>Xpsr549-2B</i> (21)	PSR549	(1,3A)
<i>Xpsr566-2A,D</i> (21)	PSR566	
<i>Xpsr593-2B</i> (21)	PSR593	(4,7B)

Table 1. Continued.

Locus [synonym(s)]	DNA clone(s)	Locations of other loci detected with clones
<i>Xpsr649-2A,D</i> (21)	PSR649	
<i>Xpsr666-2A,B,D</i> (21)	PSR666	
<i>Xpsr899-2B</i> (21)	ABA7 (34)	(6A,D)
<i>Xpsr900-2A,B,D</i> (21)	PSR900	
<i>Xpsr903-2D</i> (19, 21)	PSR903	(3A,B,D,5D)
<i>Xpsr908-2A,D</i> (21)	PSR908	(1,6B)
<i>Xpsr912-2A,B,D</i> (21)	PSR912	(5A,D)
<i>Xpsr928-2A,D</i> (21)	PSR928	
<i>Xpsr933-2A,D</i> (21)	PSR933	
<i>Xpsr946-2D</i> (21)	PSR946	(7A,DL,DS)
<i>XRbcs-2A,B,D</i> (12, 29) [<i>rbcs-2A,B,D</i> (29)]	pW9(9), pTS512 (59)	
<i>XRbpcα-2A,B,D</i> (29)	pSV10 (29)	
<i>XSbp-2B(1)</i> (19, 20)	S9.2 (25)	(2BL,3A,B,D,7B)
<i>XSs2-2A,B,D</i> (21)	pST3 (47)	
Group 2L		
<i>XGapd1-2A,B,D</i> (12)	pZm57 (8)	
<i>XksuD22-2D</i> (32)	pTtksuD22	
<i>XksuD23-2D</i> (32)	pTtksuD23	
<i>Xpsr101-2A,B,D</i> (57)	PSR101	
<i>Xpsr102-2A,B,D</i> (21, 56)	PSR102	
<i>Xpsr112-2A,B,D</i> (21, 56)	PSR112	
<i>Xpsr151-2A,B,D</i> (21)	PSR151	
<i>Xpsr304-2A,B,D</i> (21)	PSR304	
<i>Xpsr331-2A,B,D</i> (21)	PSR331	
<i>Xpsr380-2A,B,D</i> (21)	PSR380	
<i>Xpsr388-2A,B,D</i> (21)	PSR388	
<i>Xpsr390-2A,B,D</i> (21)	PSR390	
<i>Xpsr540-2A,B,D</i> (21)	PSR540	(7B)
<i>Xpsr571-2A,B,D</i> (21)	PSR571	
<i>Xpsr609-2A,B,D</i> (21)	PSR609	
<i>Xpsr630-2A,B,D</i> (21)	PSR630	
<i>Xpsr641-2A,B,D</i> (21)	PSR641	
<i>Xpsr681-2A,B,D</i> (21)	PSR681	(6D,7B)
<i>Xpsr687-2A,B,D</i> (21)	PSR687	(7A,B,D)
<i>Xpsr692-2A,B,D</i> (21)	PSR692	
<i>Xpsr901-2A,B,D</i> (21)	PSR901	
<i>Xpsr919-2A,B,D</i> (21)	PSR919	
<i>Xpsr932-2A,B,D</i> (21)	PSR932	
<i>Xpsr934-2A,B,D</i> (21)	PSR934	
<i>XSbp-2B(2)</i> (20, 21)	S9.2 (25)	(2BS,3A,B,D,7B)
Group 2		
<i>Xglk76-2A,B</i> (45) [<i>Xglk76a,b</i> (45)]	pTag76	
<i>Xglk94-2D</i> (45)	pTag94	(6A,1,3D)

Table 1. Continued.

Locus [synonym(s)]	DNA clone(s)	Locations of other loci detected with clones
[<i>XgIk194c</i> (45)]		
<i>XgIk175-2D</i> (45)	pTag175	
<i>XgIk222-2D</i> (45)	pTag222	
<i>XgIk293-2D</i> (45)	pTag293	
<i>XgIk302-2B</i> (45)	pTag302	(4A)
[<i>XgIk302b</i> (45)]		
<i>XgIk331-2B</i> (45)	pTag331	
<i>XgIk370-2B</i> (45)	pTag370	
<i>XgIk398-2B(1), (2)</i> (45)	pTag398	
[<i>XgIk398a,b</i> (45)]		
<i>XgIk400-2B</i> (45)	pTag400	
<i>XgIk407-2B</i> (45)	pTag407	
<i>XgIk431-2D</i> (45)	pTag431	(4B)
<i>XgIk431a</i> (45)]		
<i>XgIk452-2A</i> (45)	pTag452	(4A)
[<i>XgIk452b</i> (45)]		
<i>XgIk460-2A</i> (45)	pTag460	
<i>XgIk471-2B</i> (45)	pTag471	
[<i>XgIk471b</i> (45)]		
<i>XgIk520-2A</i> (45)	pTag520	(5A,1,3,6B)
[<i>XgIk520c</i> (45)]		
<i>XgIk529-2B,D</i> (45)	pTag529	
[<i>XgIk529a,b</i> (45)]		
<i>XgIk539-2B</i> (45)	pTag539	
<i>XgIk546-2B(1), (2)</i> (45)	pTag546	(5,7A,3,6B)
[<i>XgIk546e,f</i> (45)]		
<i>XgIk554-2A,B</i> (45)	pTag554	(5B)
[<i>XgIk554a,c</i> (45)]		
<i>XgIk578-2B</i> (45)	pTag578	(4A,B)
[<i>XgIk578b</i> (45)]		
<i>XgIk592-2B</i> (45)	pTag592	
<i>XgIk594-2B</i> (45)	pTag594	
<i>XgIk600-2A,B</i> (45)	pTag600	
[<i>XgIk600a,b</i> (45)]		
<i>XgIk605-2B</i> (45)	pTag605	
<i>XgIk609-2B,D</i> (45)	pTag609	
[<i>XgIk609b,a</i> (45)]		
<i>XgIk610-2A</i> (45)	pTag610	
[<i>XgIk610a</i> (45)]		
<i>XgIk613-2D</i> (45)	pTag613	
<i>XgIk618-2B</i> (45)	pTag618	
<i>XgIk632-2A,B</i> (45)	pTag632	
[<i>XgIk632a,b</i> (45)]		
<i>XgIk653-2A,B</i> (45)	pTag653	
[<i>XgIk653a,b</i> (45)]		
<i>XgIk661-2B</i> (45)	pTag661	(4A,B,D)
[<i>XgIk661c</i> (45)]		
<i>XgIk664-2A,B</i> (45)	pTag664	

Table 1. Continued.

Locus [synonym(s)]	DNA clone(s)	Locations of other loci detected with clones
[<i>Xglk664a,b</i> (45)]		
<i>Xglk684-2A</i> (45)	pTag684	
<i>Xglk687-2B</i> (45)	pTag687	
<i>Xglk699-2B</i> (45)	pTag699	
<i>Xglk703-2B</i> (45)	pTag703	
<i>Xglk734-2D</i> (45)	pTag734	
<i>Xglk738-2A</i> (45)	pTag738	
<i>Xglk740-2A,B</i> (45)	pTag740	
[<i>Xglk740b.a</i> (45)]		
<i>XksuA2-2D</i> (32)	pTtksuA2	
<i>XksuB3-2D</i> (32)	pTtksuB3	
<i>XksuC3-2D</i> (32)	pTtksuC3	
<i>XksuD8-2D</i> (32)	pTtksuD8	
<i>XksuD18-2D(1)</i> (32)	pTtksuD18	(<i>Ae. squarrosa</i> 4D)
[<i>XksuD18(A)-2D</i> (32)]		
<i>XksuD18-2D(2)</i> (32)	pTtksuD18	
[<i>XksuD18(B)-2D</i> (32)]		
<i>XksuF19-2D</i> (32)	pTtksuF19	(<i>Ae. squarrosa</i> 6D)
<i>XksuF41-2D</i> (32)	pTtksuF41	
<i>XksuG5-2D</i> (32)	pTtksuG5	
Group 3S		
<i>Xksu22-3A,B,D</i> (39)	pHv22	
<i>XksuB8-3D</i> (32)	pTtksuB8	
<i>XksuD19-3D</i> (32)	pTtksuD19	
<i>Xpsr123-3A,B,D</i> (57)	PSR123	
<i>Xpsr305-3A,B,D</i> (19)	PSR305	
<i>Xpsr383-3A,B,D</i> (19)	PSR383	
<i>Xpsr598-3A,B,D</i> (19)	PSR598	
<i>Xpsr689-3A,B,D</i> (21)	PSR698	
<i>Xpsr902-3A,B,D</i> (19)	PSR902	
<i>Xpsr903-3A,B,D</i> (19,21)	PSR903	(2D,5D)
<i>Xpsr907-3B</i> (19)	PSR907	
<i>Xpsr909-3A,B,D</i> (21)	PSR909	
<i>Xpsr910-3A,B,D</i> (21)	PSR910	
<i>Xpsr930-3A,B</i> (21)	PSR930	
<i>Xpsr1196-3A,B,D</i> (15,21))	PSR1196	
Group 3L		
<i>XCxp1-3A,B,D</i> (19)	pkc.3 (22)	
<i>XEmp-3B</i> (17)	pGC19 (33)	(5A,B,D, 6A,B,7D)
<i>XGlb33-3A,B,D</i> (21)	P7E (26)	
<i>XGlb35-3B,D</i> (21)	G5 (26)	
<i>Xpsr56-3A,B,D</i> (19, 21)	PSR56	(7A,B,D)
<i>Xpsr74-3A,B,D</i> (19)	PSR74	
<i>Xpsr78-3A,B,D</i> (19)	PSR78	
<i>Xpsr116-3A,B,D</i> (19)	PSR116	

Table 1. Continued.

Locus [synonym(s)]	DNA clone(s)	Locations of other loci detected with clones
<i>Xpsr125-3A,B,D</i> (19)	PSR125	
<i>Xpsr156-3A,B,D</i> (19, 57))	PSR156	
<i>Xpsr170-3A,B,D</i> (19)	PSR170	(5A,B)
<i>Xpsr347-3A,B,D</i> (19)	PSR347	
<i>Xpsr354-3A,B,D</i> (19)	PSR354	
<i>Xpsr394-3A,B,D</i> (19)	PSR394	
<i>Xpsr454-3B</i> (19)	PSR454	
<i>Xpsr543-3A,B,D</i> (19)	PSR543	
<i>Xpsr549-3A</i> (19, 21)	PSR549	(1A,2B)
<i>Xpsr570-3A,B,D</i> (19)	PSR570	
<i>Xpsr578-3A,B,D</i> (19)	PSR578	
<i>Xpsr754-3A,B,D</i> (21)	PSR754	
<i>Xpsr904-3A,D</i> (19)	PSR904	
<i>Xpsr916-3A,B,D</i> (21)	PSR916	
<i>Xpsr923-3A,B,D</i> (21)	PSR923	
<i>Xpsr931-3A,B,D</i> (21)	PSR931	
<i>Xpsr1060-3A,B,D</i> (15, 21)	PSR1060	
<i>Xpsr1067-3D</i> (15, 21)	PSR1067	
<i>Xpsr1077-3A,B,D</i> (15, 21)	PSR1077	
<i>Xpsr1149-3A,B,D</i> (15, 21)	PSR1149	
<i>Xpsr1203-3A</i> (21)	PSR1203	
<i>Xpsr1205-3A,B,D</i> (21)	PSR1205	
<i>XShp-3A,B,D</i> (19, 20)	S9.2 (25)	(2BS,BL,7B)
<i>XTlp-3A,B,D</i> (21)	pHv14 (10)	
Group 3		
<i>Xglk80-3B</i> (45)	pTag80	
<i>Xglk94-3D</i> (45)	pTag94	(6A,1,2D)
[<i>Xglk94a</i> (45)]		
<i>Xglk118-3A</i> (45)	pTag118	
<i>Xglk221-3A</i> (45)	pTag221	
<i>Xglk223-3B</i> (45)	pTag223	
<i>Xglk485-3A</i> (45)	pTag485	
<i>Xglk520-3B</i> (45)	pTag520	(2,5A,1,6B)
[<i>glk520b</i> (45)]		
<i>Xglk538-3B,D</i> (45)	pTag538	
[<i>Xglk538a,b</i> (45)]		
<i>Xglk546-3B</i> (45)	pTag546	(5,7A,2,6B)
[<i>Xglk546c</i> (45)]		
<i>Xglk577-3A</i> (45)	pTag577	
<i>Xglk595-3A</i> (45)	pTag595	(1B)
[<i>Xglk595a</i> (45)]		
<i>Xglk637-3B</i> (45)	pTag637	
<i>Xglk645-3A</i> (45)	pTag645	
<i>Xglk652-3B</i> (45)	pTag652	(1D)
[<i>Xglk652b</i> (45)]		
<i>Xglk683-3B</i> (45)	pTag683	

Table 1. Continued.

Locus [synonym(s)]	DNA clone(s)	Locations of other loci detected with clones
<i>Xgk718-3A,B</i> (45) [<i>Xgk718a,b</i> (45)]	pTag718	
<i>Xgk724-3B</i> (45) [<i>Xgk724d</i> (45)]	pTag724	(5A,6A,B,D)
<i>Xgk747-3A</i> (45) <i>Xgk756-3B</i> (45) [<i>Xgk756a</i> (45)]	pTag747 pTag756	(5,6A)
<i>XksuD4-3D</i> (32) <i>XksuD7-3D</i> (32) <i>XksuD24-3D</i> (32) <i>XksuD45-3D</i> (32) <i>XksuF28-3D</i> (32) <i>XksuG45-3D</i> (32) <i>XksuH2-3D</i> (32) <i>XksuM9L-3D</i> (32) <i>Xp300-3D</i> (32)	pTtksuD4 pTtksuD7 pTtksuD24 pTtksuD45 pTtksuF28 pTtksuG45 pTtksuH2 pTtksuM9L pP300	(<i>Ae squarrosa</i> 7D)
Group 4S		
<i>Xcn1BCD93-4A</i> (1) <i>Xcn1CDO484-4A</i> (1) <i>Xcn1CDO780-4A</i> (1) <i>XNra-4A</i> (13) [<i>XNra-4B</i> (11,13)]	BCD93 CDO484 CDO780 bNRp10 (14)	(7A,D) (5B,D) (7A,D)
<i>Xpsr110-4A</i> (43) 4B,D (28) <i>Xpsr115-4A</i> (44) <i>Xpsr119-4A</i> (13) [<i>Xpsr119-4B</i> (11, 13)]	PSR110 PSR115 PSR119	 (5B,D)
<i>Xpsr139-4A,B,D</i> (44) <i>Xpsr144-4A,B,D</i> (57) [<i>Xpsr144-B,A</i> (57)] <i>Xpsr147-4A,B,D</i> (44) <i>Xpsr153-4A,B,D</i> (44) <i>Xpsr160-4A</i> (13) [<i>Xpsr160-4B</i> (11,13)]	PSR139 PSR144 PSR147 PSR153 PSR160	 (5A,B,D)
<i>Xpsr166-4A,B,D</i> (44) <i>Xpsr580-4A</i> (44) <i>Xpsr593-4B</i> (21) <i>XWx-4A</i> (13) [<i>XWx-4B</i> (11,13)]	PSR166 PSR580 PSR593 pcwx27 (55)	 (5B,D) (2B,7B)
Group 4L		
<i>Xβ-Amy-B1,D1</i> (58) [<i>B1: PSR1-4A</i> (58), <i>Xβ-Amy-A1</i> (49); <i>D1. PSR1-4D</i> (58)] <i>Xcn1BCD1302-4B,D</i> (1)	pcβc51 (41) BCD1302	 (5A)

Table 1. Continued.

Locus [synonym(s)]	DNA clone(s)	Locations of other loci detected with clones
<i>Xcn1CDO1312-4B,D</i> (1)	CDO1312	(1B,5A)
<i>Xcn1WG114-4B,D</i> (1)	WG114	(5A)
<i>XFbp-4A,B,D</i> (12)	F16 (52)	
<i>Xpsr104-4A</i> (43) <i>4B,D</i> (28)	PSR 104	
<i>Xpsr163-4A,B,D</i> (57)	PSR 163	
[<i>Xpsr163-4B,A</i> (57)]		
<i>Xpsr164-4B,D</i> (44)	PSR 164	(5A)
Group 4		
<i>Xglk128-4A</i> (45)	pTag128	
<i>Xglk167-4A</i> (45)	pTag167	
<i>Xglk210-4A</i> (45)	pTag210	
<i>Xglk300-4B</i> (45)	pTag300	
<i>Xglk302-4A</i> (45)	pTag302	(2B)
[<i>Xglk302a</i> (45)]		
<i>Xglk315-4A</i> (45)	pTag315	
<i>Xglk335-4B</i> (45)	pTag335	
<i>Xglk348-4D</i> (45)	pTag348	
[<i>Xglk348a</i> (45)]		
<i>Xglk354-4A</i> (45)	pTag354	(5B)
[<i>Xglk354a</i> (45)]		
<i>Xglk431-4B</i> (45)	pTag431	(2D)
[<i>Xglk431b</i> (45)]		
<i>Xglk450-4A</i> (45)	pTag450	
<i>Xglk452-4A</i> (45)	pTag452	(2A)
[<i>Xglk452a</i> (45)]		
<i>Xglk512-4A</i> (45)	pTag512	(6A)
<i>Xglk556-4B</i> (45)	pTag556	
<i>Xglk578-4A,B</i> (45)	pTag578	(2B)
[<i>Xglk578a,c</i> (45)]		
<i>Xglk619-4A</i> (45)	pTag619	
<i>Xglk650-4A</i> (45)	pTag650	
<i>Xglk661-4A(1),(2)</i> <i>4B,D</i> (45)	pTag661	(2B)
[<i>Xglk694b,a</i> (45)]		
<i>Xglk694-4A,B</i> (45)	pTag694	
[<i>Xglk661a,e,d,b</i> (45)]		
<i>Xglk708-4A</i> (45)	pTag708	
<i>Xglk752-4A</i> (45)	pTag752	(6B)
[<i>Xglk752a</i> (45)]		
<i>XksuB5-4D</i> (32)	pTtksuB5	
<i>XksuC2-4D</i> (32)	pTtksuC2	
<i>XksuD21-4D</i> (32)	pTtksuD21	
<i>XksuF43-4D(1)</i> (32)	pTtksuF43	(5D, Ae. squarrosa)
[<i>XksuF43(A)-4D</i> (32)]		(6D)
<i>XksuF43-4D(2)</i> (32)	pTtksuF43	(5D)
[<i>XksuF43(B)-4D</i> (32)]		

Table 1. Continued.

Locus [synonym(s)]	DNA clone(s)	Locations of other loci detected with clones
Group 5S		
<i>XAc11-5A,B,D</i> (18)	pACP11	
<i>Xα-Amy-5A,B,D</i> [α Amy3 (7)]	Amy33	
<i>Nor-D3</i> (2,23,48)	Derivatives of pTa250 (30)	
<i>Xpsr118-5A,B,D</i> (57)	PSR118	
<i>Xpsr170-5A,B</i> (19)	PSR170	(3A,B,D)
<i>Xpsr903-5D</i> (19,21)	PSR903	(2D,3A,B,D)
<i>5S-Rrna-A,B2</i> (24) [5SDna-A,B2 (24)]	pTa794 (31)	(5D)
Group 5L		
<i>XAc13-5B</i> (18)	pACP1	(7A,B,D)
β - <i>Amy-A1</i> (58) [PSR1-5A (58) <i>Xβ-Amy-B1</i> (49)]	pc β C51 (41)	
<i>Xcn1BCD87-5B,D</i> (1)	BCD87	(7B)
<i>Xcn1BCD1302-5A</i> (1)	BCD1302	(4B,D)
<i>Xcn1CDO484-5B,D</i> (1)	CDO484	(4A)
<i>Xcn1CDO1312-5A</i> (1)	CDO1312	(1B,4B,D)
<i>Xcn1WG114-5A</i> (1)	WG114	(4B,D)
<i>XEmbp-5A,B,D</i> (17)	pGC19 (33)	(3B,6A,B,7D)
<i>Xksu8-5A,B,D</i> (39)	pHv8, pHv75	
<i>Xksu24-5A,B,D</i> (1),(2) (39) [<i>Xksu24-5(1)A,B,D</i> (39)]	pHv24	
<i>Xksu26-5A,B,D</i> (39)	pHv26, pHv29	
<i>Xksu58-5A,B,D</i> (39)	pHv58	
<i>Xpsr79-5A,B,D</i> (44) [<i>Xpsr81</i> (44)]	PSR79	
<i>Xpsr100-5A,B,D</i> (21)	PSR100	(2A,B,D)
<i>Xpsr109-5A,B,D</i> (21)	PSR109	(2A,B,D)
<i>Xpsr115-5B,D</i> (44)	PSR115	(4A)
<i>Xpsr120-5A,B,D</i> (1),(2), (3) (44)	PSR120	
<i>Xpsr128-5A,B,D</i> (57)	PSR128	
<i>Xpsr145-5A,B,D</i> (44)	PSR145	
<i>Xpsr147-5A,B,D</i> (28)	PSR147	(4A,B,D)
<i>Xpsr150-5A,B,D</i> (44)	PSR150	(2A,B,D,7A,B,D)
<i>Xpsr164-5A</i> (44)	PSR164	(4A,B)
<i>Xpsr360-5A,B,D</i> (44)	PSR360	
<i>Xpsr426-5A,B,D</i> (44)	PSR426	
<i>Xpsr580-5B,D</i> (44)	PSR580	(4A)
<i>Xpsr912-5A,D</i> (21)	PSR912	(2A,B,D)
<i>XRbcs-5A,B,D</i> (29) [<i>rbc5-5A,B,D</i> (29)]	pTS512 (59)	

Table 1. Continued.

Locus [synonym(s)]	DNA clone(s)	Locations of other loci detected with clones
Group 5		
<i>Xgk83-5B</i> (45)	pTag83	
<i>Xgk157-5D</i> (45)	pTag157	
<i>Xgk165-5B</i> (45)	pTag165	
<i>Xgk251-5D</i> (45)	pTag251	
<i>Xgk278-5A,B</i> (45)	pTag 278	(6B)
[<i>Xgk278a,b</i> (45)]		
<i>Xgk317-5A</i> (1), (2) (45)	pTag317	(6A)
[<i>Xgk317a,b</i> (45)]		
<i>Xgk319-5B</i> (45)	pTag319	
<i>Xgk354-5B</i> (45)	pTag354	(4A)
[<i>Xgk354b</i> (45)]		
<i>Xgk424-5A</i> (45)	pTag424	
<i>Xgk505-5A</i> (45)	pTag505	
<i>Xgk510-5A,B</i> (45)	pTag510	
[<i>Xgk510a,b</i> (45)]		
<i>Xgk520-5A</i> (45)	pTag520	(2A,1,3,6B)
[<i>Xgk520e</i> (45)]		
<i>Xgk546-5A</i> (1),(2)(45)	pTag546	(7A,2,3,6B)
[<i>Xgk546a,g</i> (45)]		
<i>Xgk554-5B</i> (45)	pTag554	(2A,B)
[<i>Xgk554b</i> (45)]		
<i>Xgk587-5A,D</i> (45)]	pTag587	
[<i>Xgk587a,b</i> , (45)]		
<i>Xgk612-5A</i> (45)	pTag612	
<i>Xgk614-5A</i> (45)	pTag614	
<i>Xgk621-5D</i> (45)	pTag621	
<i>Xgk629-5B</i> (45)	pTag629	
<i>Xgk644-5A</i> (45)	pTag644	
<i>Xgk695-5D</i> (45)	pTag695	
<i>Xgk701-5A</i> (45)	pTag701	
<i>Xgk724-5A</i> (45)	pTag724	(6A,B,D,3B)
[<i>Xgk724a</i> (45)]		
<i>Xgk756-5A</i> (45)	pTag756	(6A,3B)
[<i>Xgk756c</i> (45)]		
<i>XksuD42-5D</i> (32)	pTksuD42	
<i>XksuF43-5D</i> (32)	pTksuF43	(4D, <i>Ae. squarrosa</i> 6D)
[<i>XksuF43(A)-5D</i> (32)]		
<i>XksuF43-5D</i> (2) (32)	pTksuF43	(4D)
[<i>XksuF43 (B)-5D</i> (32)]		
<i>XksuM2-5D</i> (32)	pTksuM2	
<i>XksuM4-5D</i> (32)	pTksuM4	
<i>XksuM70-5D</i> (32)	pTksuM70	
<i>Xpsr170-5B</i> (19)	PSR170	(5AS,3A,B)
<i>5S-Rrna-D2</i> (24)	pTa794 (31)	(5AS,5BS)
[<i>5S-Dna-D2</i> (24)]		

Table 1. Continued.

Locus [synonym(s)]	DNA clone(s)	Locations of other loci detected with clones
Group 6S		
<i>XCxp-6A,B,D</i> (6)	2437 (5)	
<i>XEmbp-6B</i> (17)	pGC19 (33)	(6AL,3B,5A,B,D,7D)
<i>XGli-6A,B,D</i> (36)	pTag53	
<i>XksuG44-6D</i> (32)	pTtksuG44	(<i>Ae. squarrosa</i> 5D)
<i>XksuH3-6D</i> (32)	pTtksuH3	
<i>XNra-6A,B,D</i> (40)	bNRp10 (14)	
[<i>XNar-6A,B,D</i> (40)]		
<i>Nor-B2</i> (2,48)	Derivatives of pTa250 (30)	
<i>XPgk2-6A,B,D</i> (12)	p20 (46)	
<i>Xpsr167-6A,B,D</i> (57)	PSR167	
<i>Xpsr681-6D</i> (21)	PSR681	(2A,B,D,7B)
<i>Xpsr899-6A,D</i> (21)	ABA7 (34)	(2B)
<i>Xpsr904-6A</i> (19)	PSR904	(3A,D)
Group 6L		
<i>Xα-Amy-6A,B,D</i> (42)	2119	
[<i>a-Amy1</i> (42)]		
<i>XEmbp-6A</i> (17)	pGC19 (33)	(6BS,3B,5A,B,D,7D)
<i>XksuD17-6D</i> (32)	pTtksuD17	
<i>XPrk-6A,B,D</i> (12)	F6 (53)	
<i>Xpsr154-6A,B,D</i> (57)	PSR154	
<i>Xpsr908-6B</i> (21)	PSR908	(1B,2A,D)
Group 6		
<i>Xglk94-6A</i> (45)	pTag94	(1,2,3D)
[<i>Xglk94b</i> (45)]		
<i>Xglk172-6A</i> (45)	pTag172	(7A,B)
[<i>Xglk172a</i> (45)]		
<i>Xglk229-6B</i> (45)	pTag229	
<i>Xglk259-6A</i> (45)	pTag259	
<i>Xglk299-6A,D</i> (45)	pTag299	
[<i>Xglk299a,b</i> (45)]		
<i>Xglk317-6A</i> (45)	pTag317	(5A)
[<i>Xglk317c</i> (45)]		
<i>Xglk334-6A</i> (45)	pTag 334	
<i>Xglk479-6A</i> (45)	pTag479	
<i>Xglk495-6D</i> (45)	pTag495	
<i>Xglk512-6A</i> (45)	pTag512	(4A)
[<i>Xglk512a</i> (45)]		
<i>Xglk520-6B</i> (45)	pTag520	(2,5A,1,3B)
[<i>Xglk520d</i> (45)]		
<i>Xglk537-6A</i> (45)	pTag537	
<i>Xglk546-6B</i> (45)	pTag546	(5,7A,2,3B)
[<i>Xglk546b</i> (45)]		

Table 1. Continued.

Locus [synonym(s)]	DNA clone(s)	Locations of other loci detected with clones
<i>Xgk547-6A</i> (1),(2),(3), <i>6B</i> (45) [<i>Xgk547a,b,d,c</i> (45)]		
<i>Xgk562-6A</i> (45)	pTag562	
<i>Xgk582-6B</i> (45)	pTag582	
<i>Xgk680-6B</i> (45)	pTag680	
<i>Xgk705-6B</i> (45)	pTag705	
<i>Xgk724-6A,B,D</i> (45) [<i>Xgk724e,c,b</i> (45)]	pTag724	(5A,3B)
<i>Xgk736-6B</i> (45)	pTag736	
<i>Xgk744-6B</i> (45)	pTag744	
<i>Xgk752-6B</i> (45) [<i>Xgk752b</i> (45)]	pTag572	(4A)
<i>Xgk756-6A</i> (45) [<i>Xgk756b</i> (45)]	pTag756	(5A,3B)
<i>Xgk762-6A</i> (45)	pTag762	
<i>XksuB6-6D</i> (32)	pTtksuB6	
<i>XksuD1-6D</i> (32)	pTtksuD1	
<i>XksuD11-6D</i> (32)	pTtksuD11	
<i>XksuF24-6D</i> (32)	pTtksuF24	(7D)
<i>XksuM9S-6D</i> (32)	pTtksuM9S	(<i>Ae. squarrosa</i> 5D)
Group 7S		
<i>Xacl3-7A,B,D</i> (18)	pACP1(35)	(5B)
<i>XcnlBCD87-7B</i> (1)	BCD87	(5B,D)
<i>XcnlBCD93-7A,D</i> (1)	BCD93	(4A)
<i>XcnlCDO780-7A,D</i> (1)	CDO780	(4A)
<i>XksuA1-1D</i> (32)	pTtksuA1	
<i>XNra-7A,D</i> (13)	bNRp10(14)	
<i>Xpsr65-7A,B,D</i> (11,13)	PSR65	
<i>Xpsr103-7A,B,D</i> (11,13)	PSR103	
<i>Xpsr108-7A,B,D</i> (21)	PSR108	(2A,B,D)
<i>Xpsr119-7A,D</i> (13)	PSR119	
<i>Xpsr150-7A,B,D</i> (21)	PSR150	(2A,B,D,5A,B,D)
<i>Xpsr152-7A,B,D</i> (11,13,57)	PSR152	
<i>Xpsr160-7A,D</i> (11,13)	PSR160	
<i>Xpsr540-7B</i> (21)	PSR540	(2A,B,D)
<i>Xpsr946-7D(1)</i> (21)	PSR946	(2D,7AL,7DL)
<i>XSs1-7A,B,D</i> (47) [<i>Ss1</i> (47)]	pST8	
<i>XWx-7A,D</i> (13)	pcwx27(55)	
Group 7L		
<i>Xα-Amy-7A,B,D</i> (11,13,42) [<i>α-Amy2</i> (42)]	4868(42), <i>Amy2/46</i> (38)	
<i>XEmbp-7D</i> (17)	pGC19(33)	(3B,5A,B,D,6A,B)
<i>XFed-7A,B,D</i> (16)	1.3 Kb <i>Hind</i> III fragment of a wheat gene	

Table 1. Continued.

Locus [synonym(s)]	DNA clone(s)	Locations of other loci detected with clones
<i>XGapd2-7A,B,D</i> (12)	pZm9 (8)	
<i>XksuA5-7D</i> (32)	pTtksuA5	
<i>XPepc-7A,B,D</i> (13,32)	SORCO (64)	
<i>Xpsr56-7A,B,D</i> (11,13,19)	PSR56	(3A,B,D)
[<i>Xpsr117</i> (11,13)]		
<i>Xpsr72-7A,B,D</i> (11,13)	PSR72	
<i>Xpsr105-7A,B,D</i> (11,13)	PSR105	
<i>Xpsr121-7A,B,D</i> (11,13)	PSR121	(1A,B,D)
<i>Xpsr129-7A,B,D</i> , (11,13,57)	PSR129	
<i>Xpsr165-7A,B,D</i> (11,13)	PSR165	
<i>Xpsr169-7A,B,D</i> (11,13)	PSR169	
<i>Xpsr593-7B</i> (21)	PSR593	(2,4B)
<i>Xpsr681-7B</i> (21)	PSR681	(2A,B,D,6D)
<i>Xpsr687-7A,B,D</i> (21)	PSR687	(2A,B,D)
<i>Xpsr946-7A,7D(2)</i> (21)	PSR946	(2A,7DS)
<i>XSbp-7B</i> (19,20)	S9.2 (25)	(2B,3A,B,D)
Group 7		
<i>Xglk35-7A,B</i> (45)	pTag35	
[<i>Xglk35b,a</i> (45)]		
<i>Xglk61-7B</i> (45)	pTag61	
<i>Xglk172-7A,B</i> (45)	pTag172	(6A)
[<i>Xglk172b,c</i> (45)]		
<i>Xglk184-7D(1),(2)</i> (45)	pTag184	
[<i>Xglk184a,b</i> (45)]		
<i>Xglk197-7B</i> (45)	pTag197	
<i>Xglk301-7A</i> (45)	pTag301	
<i>Xglk341-7A,D</i> (45)	pTag341	
[<i>Xglk341b,a</i> (45)]		
<i>Xglk349-7B</i> (45)	pTag349	
<i>Xglk356-7B</i> (45)	pTag356	
<i>Xglk439-7B</i> (45)	pTag439	
<i>Xglk478-7B</i> (45)	pTag478	
<i>Xglk536-7B</i> (45)	pTag536	
<i>Xglk546-7A</i> (45)	pTag546	(5A,2,3,6B)
[<i>Xglk546d</i> (45)]		
<i>Xglk549-7B</i> (45)	pTag549	(1B)
[<i>Xglk549a</i> (45)]		
<i>Xglk576-7A</i> (45)	pTag576	
<i>Xglk598-7B</i> (45)	pTag598	
<i>Xglk642-7A</i> (45)	pTag642	
<i>Xglk651-7A</i> (45)	pTag651	
<i>Xglk658-7A</i> (45)	pTag658	
[<i>Xglk658a</i> (45)]		
<i>Xglk686-7A</i> (45)	pTag686	
<i>Xglk702-7D</i> (45)	pTag702	
<i>Xglk750-7B</i> (45)	pTag750	
<i>XksuB7-7D</i> (32)	pTtksuB7	
<i>XksuD6-7D</i> (32)	pTtksuD6	

Table 1 Continued

Locus [synonym(s)]	DNA clone(s)	Locations of other loci detected with clones
<i>XksuD10-7D</i> (32)	pTtksuD10	
<i>XksuD25-7D</i> (32)	pTtksuD25	
<i>XksuD46-7D</i> (32)	pTtksuD46	
<i>XksuE7-7D</i> (32)	pTtksuE7	
<i>XksuF2-7D</i> (1),(2), (4),(5) (32) [<i>XksuF2</i> (A),(B),(D) (E)-7D (32)]	pTtksuF2	(<i>Ae squarrosa</i> 2D)
<i>XksuF2-7D</i> (3) (32) [<i>XksuF2</i> (C)-7D (32)]	pTtksuF2	
<i>XksuF24-7D</i> (3) (32)	pTtksuF24	(6D)

^a S designates arm groups composed of short arms and L designates arm groups composed of long arms except in group 4 where S designates 4AL, 4BS, and 4DS, and L designates 4AS, 4BL, and 4DL

^b See the text for an explanation of RFLP locus nomenclature. In this table, the expression 'A,B,D1' is an abbreviation for A1, B1, and D1 and the expression '1A,B,D' is an abbreviation for 1A, 1B, and 1D. Symbols used in locus designations are explained below.

<i>X</i>	Basic symbol for all RFLP loci except known protein and RNA loci	<i>Glb3</i>	(1-3)- β -glucanase (EC3 2 1 39)
<i>cnl</i>	Cornell University, Ithaca, New York, U S A	<i>Gli</i>	Gliadin
<i>glk</i>	Genetics Laboratory, Kyoto University, Kyoto, Japan	<i>Glu</i>	Glutenin
<i>ksu</i>	Kansas State University, Manhattan, Kansas, U S A	<i>Ica</i>	Chymotrypsin inhibitor
<i>psr</i>	Institute for Plant Science Research, Cambridge Laboratory, Norwich, U K	<i>Lec</i>	Lectin
<i>Acl</i>	Acyl carrier protein	<i>Nor</i>	Nucleolus organizer region
<i>Adh</i>	Alcohol dehydrogenase	<i>Nra</i>	Nitrate reductase
α - <i>Amy</i>	α -Amylase	<i>Pepc</i>	Phosphoenolpyruvate carboxylase
β - <i>Amy</i>	β -Amylase	<i>Per</i>	Peroxidase
<i>Cxp</i>	Carboxypeptidase	<i>Pgk1</i>	Chloroplast phosphoglycerate kinase
<i>Em</i>	Early-methionine labeled polypeptide	<i>Pgk2</i>	Cytosolic phosphoglycerate kinase
<i>Embp</i>	DNA binding protein of the basic-leucine zipper class, β -ZIP	<i>Ppdk</i>	Pyruvate orthophosphate dikinase
<i>Fbp</i>	Fructose-1,6-bisphosphate	<i>Prk</i>	Phosphoribulokinase
<i>Fed</i>	Ferredoxin	<i>Rbcs</i>	Ribulose-1,5-bisphosphate carboxylase small subunit
<i>Gapd1</i>	Chloroplast glyceraldehyde phosphate dehydrogenase	<i>Rbpa</i>	Rubisco binding protein
<i>Gapd2</i>	Cytosolic glyceraldehyde phosphate dehydrogenase	<i>5S-Rrna</i>	5S ribosomal RNA
		<i>Ss</i>	Sucrose synthase
		<i>Sbp</i>	Sedoheptulose-1,7-bisphosphatase
		<i>Tlp</i>	Thiolprotease
		<i>Wx</i>	Waxy

^c References

- | | | |
|--------------------------------|--------------------------|-----------------------------|
| 1 Anderson et al (1992) | 8 Brinkmann et al (1988) | 16 Chnoy et al (1991) |
| 2 Appels and Dvorak (1982) | 9 Broglie et al (1983) | 17 Devos et al (1991a) |
| 3 Appels et al (1980) | 10 Chandler et al (1984) | 18 Devos et al (1991b) |
| 4 Bartels et al (1986) | 11 Chao et al (1988) | 19 Devos et al (1992a) |
| 5 Baulcombe and Bufford (1983) | 12 Chao et al (1989a) | 20 Devos et al (1992b) |
| 6 Baulcombe et al (1987a) | 13 Chao et al (1989b) | 21 Devos et al (1993a,b) |
| 7 Baulcombe et al (1987b) | 14 Cheng et al (1986) | 22 Doan and Fincher (1988) |
| | 15 Cheung et al (1992) | 23 Dvorak and Appels (1986) |

Table 1 Continued

24 Dvorak et al (1989)	38 Huttley et al (1988)	53 Raines et al (1989)
25 T Dyer (pers comm cited in reference 19)	39 Kam-Morgan et al (1989)	54 Rebmann et al (1991)
26 Fincher, G (pers comm)	40 Kleinhofs et al (1988)	55 Rohde et al (1988)
27 Futers et al (1990)	41 Kreis et al (1987)	56 Sharp and Soltes-Rak (1988)
28 M D Gale (pers comm)	42 Lazarus et al (1985)	57 Sharp et al (1988)
29 Gahli et al (1991)	43 Liu and Gale (1991)	58 Sharp et al (1989)
30 Gerlach and Bedbrook (1979)	44 Liu et al (1992)	59 Smith et al (1983)
31 Gerlach and Dyer (1980)	45 Liu and Tsunewaki (1991)	60 Snape et al (1985)
32 Gill et al (1991)	46 Longstaff et al (1989)	61 Thompson et al (1983)
33 Gultnan et al (1990)	47 Marana et al (1988)	62 Wang et al (1991)
34 M Gulli (pers comm)	48 May and Appels (1987)	63 Wang et al (1992)
35 Hansen (1987)	49 McIntosh et al (1989)	64 Westhoff, P (pers comm cited in reference 11)
36 Harberd et al (1985)	50 Mitchell et al (1989)	65 Williamson et al (1988)
37 Harberd et al (1986)	51 Raikhel and Wilkins (1987)	
	52 Raines et al (1988)	

determined the chromosomal locations of over 400 DNA fragments but they reported the locations of only the 197 RFLP loci for which they had F_2 segregation data; the locations of 228 fragments for which segregation data were not obtained were not reported and are not shown in Table 1. In addition, the chromosomal locations of approximately 200 wheat RFLP loci identified with 71 Texas A&M University (= *tam*) probes by Michael Devey and the author have not been published and are not included in the table. Furthermore, most of the loci on maps of several wheat chromosomes soon to be published by researchers at the Cambridge Laboratory in Norwich, United Kingdom, have not been reported to date (Michael Gale, pers. comm.) and are not shown in Table 1.

The clones listed in Table 1 are available to researchers for academic purposes almost without exception. Many of the clones are not available for commercial purposes, however, and likewise clones that have been developed but not published may not be available. Requests for clones should be directed to the laboratory at which the clones were developed and the intended use of the clones stated.

4. RFLP maps

RFLP-based genetic maps of wheat chromosomes have been reported by Chao et al. (1989b), Liu and Tsunewaki (1991), Liu et al. (1992), and Devos et al. (1992a, 1993a,c). The maps reported in the most recent of these papers (Devos et al. 1993a,c) were not available for use in preparing this review but the RFLP loci reported in the paper are listed in Table 1. The other maps are summarized in Figs. 1–6.

Both the difficulty in using the genetic variation present among wheat varieties and the value of using alien genetic variation in constructing RFLP maps of wheat chromosomes are illustrated by the findings of Chao et al. (1989b). These investigators used six different mapping populations derived

from ten varieties in constructing maps of chromosomes 7A, 7B, and 7D. They studied two F₂ populations obtained from varietal crosses, one population of single-seed descent lines derived from a varietal cross, one population of doubled haploids derived from a cross of two intervarietal chromosome substitution lines and two populations of single-chromosome recombinant lines. Thirty-one RFLP loci were mapped, including five loci in 7A, 13 in 7B, and 13 in 7D. A test of the level of polymorphism among all pair-wise combinations of six of the 10 varieties using 18 group 7 cDNA clones and 13 restriction enzymes revealed polymorphism in an average of only 8.7% of the comparisons. A major difference was observed between chromosomes, namely, 16.1% polymorphism for 7B and only 3.5% and 4.7% for 7A and 7D, respectively. However, a similar test for chromosome 7D polymorphism between Hobbit 'S' and *VPM1*, which contains a 7D that is mostly derived from *Aegilops ventricosa*, revealed 23.3% polymorphism. Similar findings were reported in a study of the homoeologous group 3 chromosomes (Devos et al. 1992a) in which only two F₂ mapping populations were used. One of the populations came from a very wide cross, namely, Chinese Spring × Synthetic; the latter, as the name implies, is a synthetic hexaploid variety derived from a cross between *T. dicoccum* (genomes A and B) and *T. tauschii* (D). The level of polymorphism in this population was more than twice that in the second population, which was derived from a varietal cross. D-genome polymorphism was only slightly more than half that of the A and B genomes, which had a similar level of polymorphism.

The chromosome maps developed by Liu and Tsunewaki (1991) are based on analysis of an F₂ population derived from a cross of Chinese Spring × *T. spelta* var. *duhamelianum*. In a test conducted with five restriction enzymes, an average of 22.7% of the clones detected one or more RFLPs and one or more RFLPs were detected with at least one of the five enzymes by 38.9% of the clones. The number of polymorphic loci was markedly less in the D-genome than in the A- and B-genomes. Only 20.9% of the fragments localized to D-genome chromosomes were polymorphic while 42.2% and 38.8% of the fragments located in the A and B genomes, respectively, were polymorphic. Only small differences were found among restriction enzymes in the level of RFLP detected, with *HindIII* revealing the lowest level and *BamHI* the highest. With the former, 20.6% and with the latter, 25.6% of the clones tested detected one or more RFLPs. The range in polymorphism found by Chao et al. (1989b) with the same five enzymes in their study of six varieties was only slightly larger. However, they also identified several enzymes that revealed a much lower level of RFLP or, in the case of *KpnI*, no RFLP among the six varieties.

The degree of polymorphism detected between clones in the aforementioned studies varied greatly. For example, in the analysis of six varieties by Chao et al. (1989a) described above, PSR clone 160 detected 20.6%, 64.0% and 49.3% polymorphism in 7A, 7B and 7D, respectively, while clones 105, 108, and 169 did not detect any polymorphism in these chromosomes. Also, clone 103 detected 0%, 42.2% and 1.4% polymorphism in 7A, 7B and 7D, respectively, clones 129, 152, and 165 detected from 2.8% to 6.1% polymorphism in 7B and no

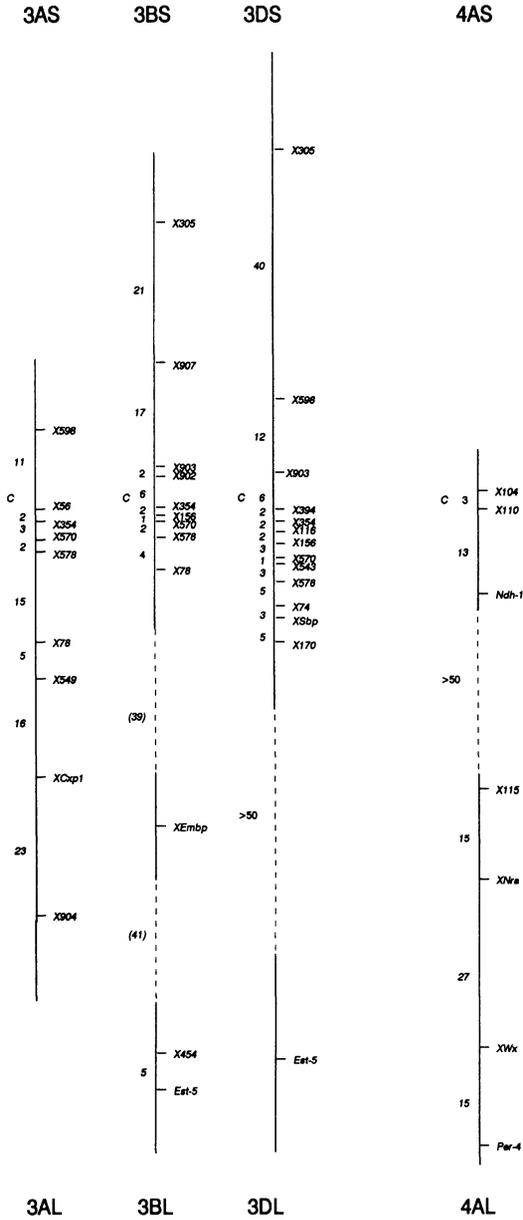
polymorphism in 7A and 7D, and clone 119 detected 40.2%, 0%, and 29.2% polymorphism in 7A, 7B and 7D, respectively.

Devos et al. (1992a) found gDNA clones to be almost twice as efficient as cDNA clones in detecting RFLPs and also that some classes of gDNA clones, namely, chromosome-specific clones and clones that hybridize to sequences in non-homoeologous chromosomes, are the most polymorphic.

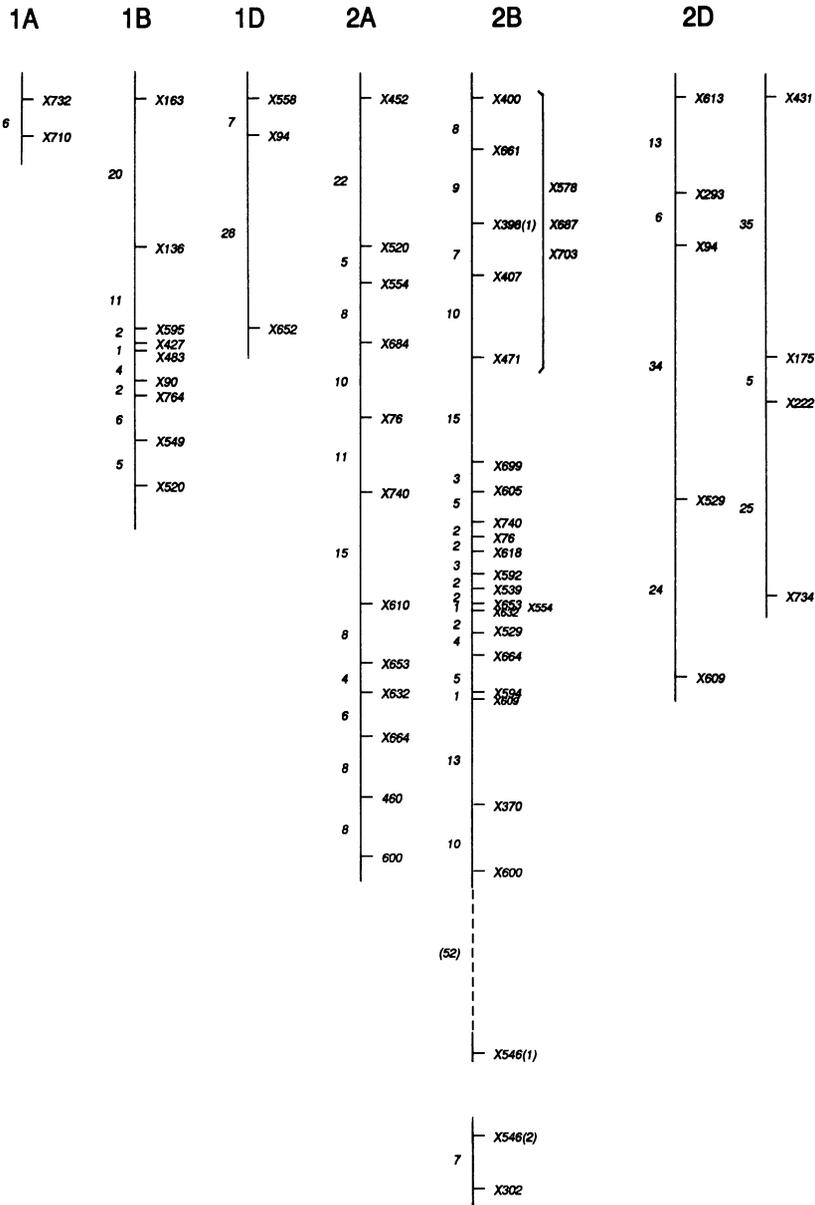
By use of ditelosomic lines, wheat loci are readily assigned to chromosome arms. This allows the positions of centromeres to be determined on chromosome maps, often quite precisely when they lie between two closely-linked markers located in different chromosome arms. A notable feature of the maps for chromosomes 3A, 3B, 3D, 7B and 7D developed by Chao et al. (1989b) and Devos et al. (1992a) (Figs. 1 and 2) is the high degree of clustering of loci around the centromeres combined with generally few and widely-spaced distal loci. It is highly probable that this is due to a much higher amount of genetic recombination in the distal as opposed to the proximal regions of wheat chromosome arms, as has been observed in the combined cytological and genetic studies of Dvorak and Chen (1984) of chromosome 6B, Snape et al. (1985) and Curtis and Lukaszewski (1991) of 1B and B.S. Gill and colleagues of the homoeologous group 7 chromosomes (B.S. Gill, pers. comm.). Clustering of loci around the centromere is not apparent on the maps of 4A and 7A; this may be due to the small number of loci mapped in these chromosomes.

Another notable feature of the aforementioned maps is the high degree of conservation of locus order among the three chromosomes of both groups 3 and 7. Conservation is complete among all of the loci that were mapped in two or more group 3 chromosomes. In the case of group 7, conservation is also complete with the exception that the order of three loci as determined for 7B in one mapping population differs from that determined for the same markers for 7B and 7D in two other mapping populations. The locus order in the first population is based on only eight recombinant gametes; this anomaly may be explained when one or more distal short-arm markers are mapped in this population. Conservation of locus order was also observed between 3R of rye and the homoeologous group 3 chromosomes of wheat and a consensus map for these four chromosomes was published (Devos et al. 1992a). Significant variation in recombination frequencies between homoeologous loci in different genomes was detected for some group 3 and group 7 loci but the available data are insufficient to indicate other than that 3D may recombine more frequently than 3A and 3B (Chao et al. 1989b; Devos et al. 1992a).

As noted above, Liu and Tsunewaki (1991) did not use ditelosomic lines in their study and consequently their maps (Figs. 3–6) do not show either the arm locations of loci or the positions of centromeres. Clusters of loci that may be indicative of a centromeric region are apparent on only a few chromosomes, e.g., 1B and 2B. A few regions that are conserved among at least two chromosomes of a homoeologous group are apparent, e.g., the region spanned by loci *X740*, *X653*, *X632*, *X664* and *X600* in chromosomes 2A and 2B. However, the vast majority of clones used to map loci in this study hybridized to a fragment located



Figs. 1-2. RFLP-based genetic maps of wheat chromosomes 3A, 3B and 3D (adapted from Devos et al. 1992a), 7A, 7B and 7D (adapted from Chao et al. 1989b and including *Per-A4* from Liu et al. 1990, and *Wsp1* from Liu and Gale 1989) and 4A (adapted from Liu et al. 1992). Locus symbols are shown to the right and genetic distances between loci in cMs to the left of the chromosomes. All X loci that are designated with a number only are *psr* loci. The approximate location of the centromere in each chromosome is indicated with a C. The maps include a small number of non-RFLP loci, e.g., *Est-B5* and *Est-D5* on 3BL and 3DL, respectively.



Figs. 3–6. RFLP-based genetic maps of the 21 chromosomes of wheat (adapted from Liu and Tsunewaki 1991). Locus symbols are shown to the right and genetic distances between loci in cMs to the left of the chromosomes. All loci, except Q (squarehead), are *glk* loci. The locations of centromeres and the arm locations of the loci, except Q, known to be in 5AL, are unknown. Unlinked loci that have been assigned to chromosomes are not shown. Two or three linkage groups are assigned to some chromosomes; the orientation of these linkage groups relative to each other is unknown.

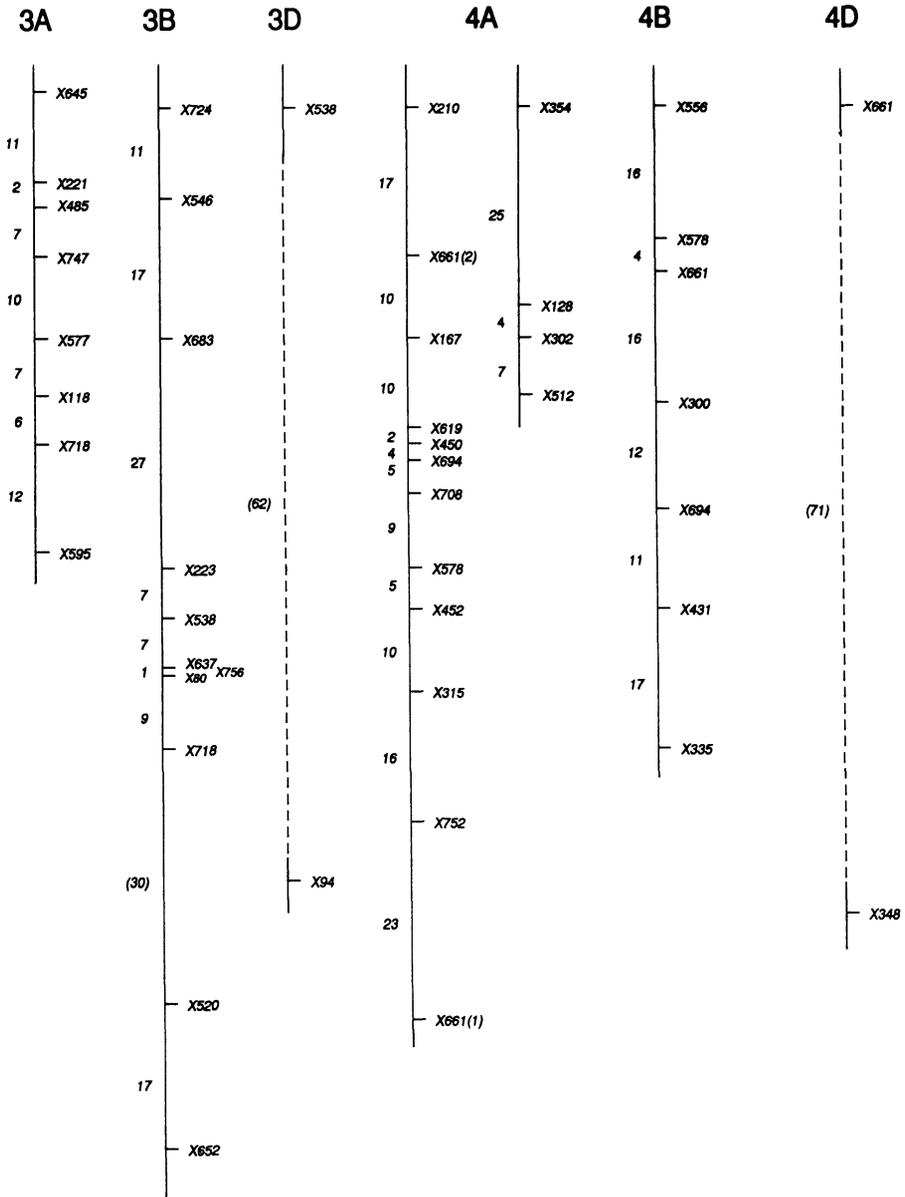


Fig. 4.

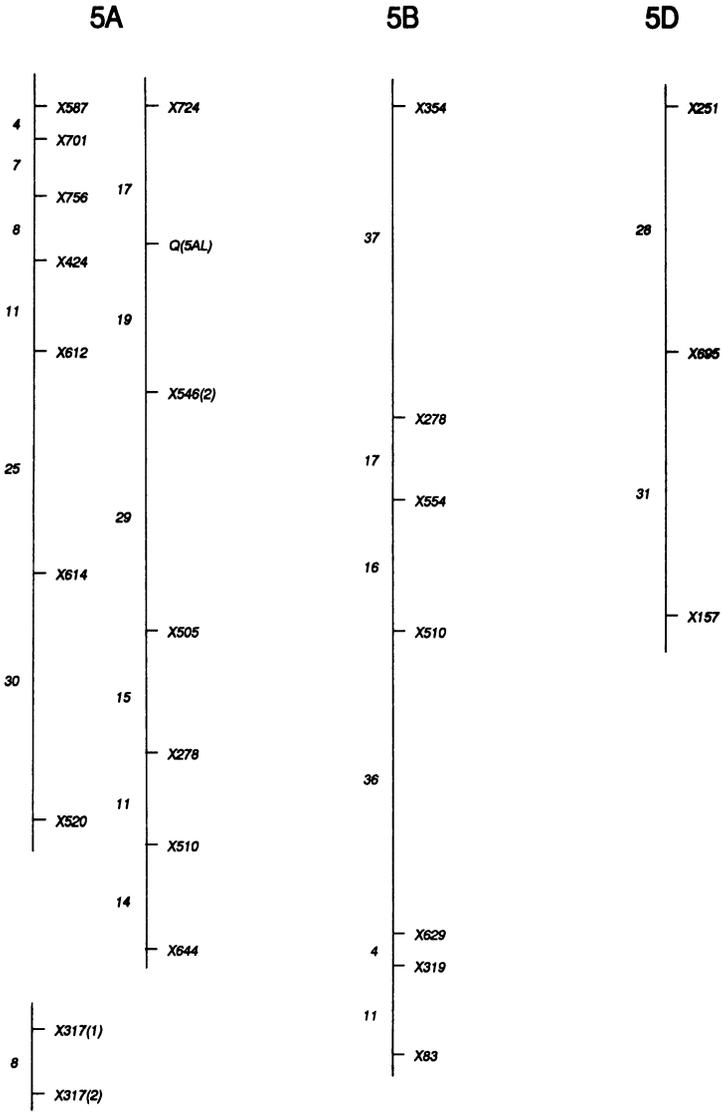


Fig. 5.

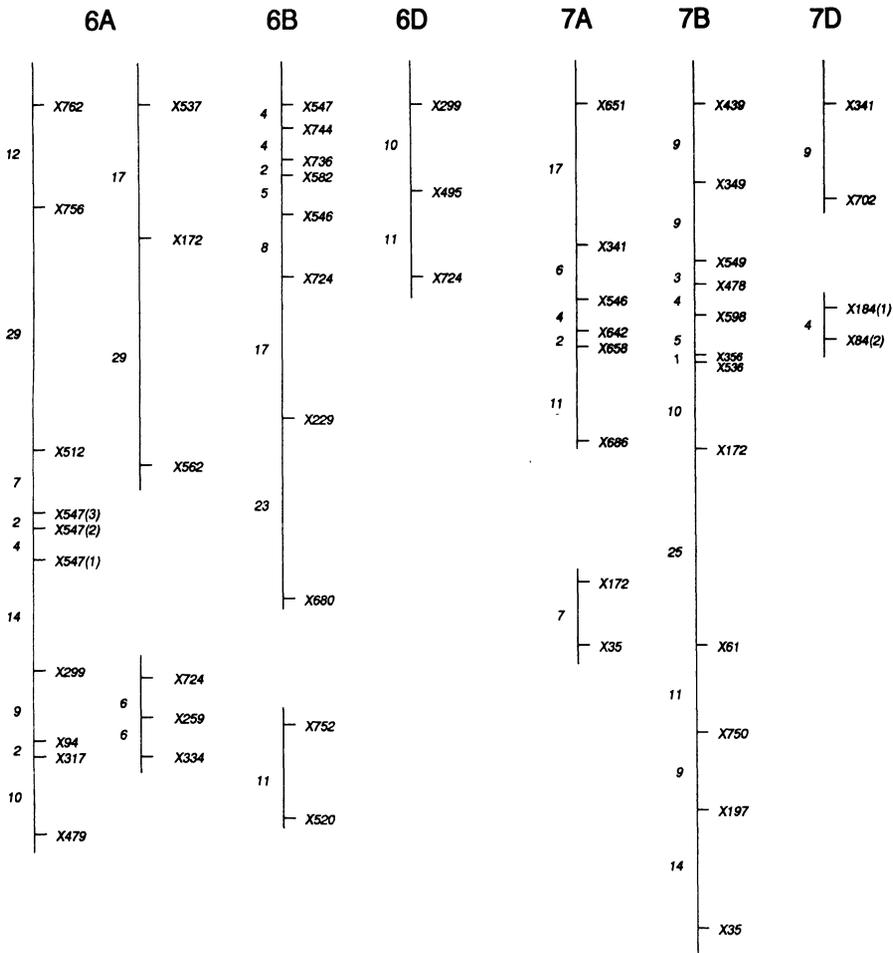


Fig. 6.

in only one chromosome (see Table 1) and thus they cannot provide information about locus order conservation. Unfortunately, no common probes were used in this study and those of groups 3 and 7 and 4A described above, so the maps cannot yet be integrated. A striking feature of the Liu and Tsunewaki (1991) maps is the very small number of loci mapped in D-genome chromosomes. As noted above, the RFLP level detected in the D-genome in the mapping population utilized was only about half of that in the A and B genomes.

5. RFLP maps of wheat relatives

Anonymous DNA clones (especially anonymous cDNA clones) derived from wheat relatives such as barley, rye and oats generally hybridize to wheat DNA fragments and they have been used and are being used to map wheat and to

develop comparative RFLP maps of these species. Findings from isozyme studies (Hart 1987) indicate that the gene synteny relationships that existed in the ancestral Triticeae genome are largely conserved in the Triticeae genomes that exist today and recent findings from RFLP studies (Devos et al. 1992a, 1993a,b) indicate a high degree of conservation of gene order among homoeologous chromosomes of Triticeae species. Consequently, there is much interest among wheat geneticists in developing maps of close relatives of wheat, especially progenitor species.

Maps of *T. tauschii* have been published by Gill et al. (1991) and by Lagudah et al. (1991). Both maps include some non-RFLP loci, e.g., isozyme loci. The former map contains 152 loci and consists of 1554 cMs. The average number of cMs per chromosome is 222 and the range is from 116 (4D) to 405 (6D). The latter map contains 58 loci and about 550 cMs. In the latter study, locus order was found to be conserved between *T. tauschii* and *T. aestivum* for the small number of loci for which linkage data was available for both species.

6. International Triticeae Mapping Initiative

In 1989, following discussions between Jan Dvorak, Bikram Gill, Cal Qualset and the author, the International Triticeae Mapping Initiative (ITMI), a consortium of individual investigators interested in mapping Triticeae genomes, was organized. Composed of wheat geneticists from several countries, the purpose of ITMI is to facilitate collaboration among investigators and maximize the rate of progress in mapping Triticeae genomes while minimizing duplication of efforts. The mapping activities have been partitioned into efforts on individual homoeologous chromosome groups and on individual genomes of diploid Triticeae species. A meeting of ITMI investigators is held each year. Scientists interested in ITMI should contact any of the member investigators or the Genetics Resources Conservation Program, University of California, Davis, through which ITMI's activities are coordinated.

7. Concluding remarks

A mere glance at Figs. 1–6 makes it apparent that a genetic map of wheat on which markers are distributed at intervals of 20 cMs or less throughout all but a small part of the genome is far from complete. Three-thousand cMs (an average of about 140 cMs per chromosome) seems a conservative estimate of the size of the wheat map and, in view of the present size of the still incomplete maps of some individual chromosomes (e.g., 7B, composed of almost 200 cMs, not including the translocated portion), a more reasonable minimum estimate would seem to be 3500 cMs (an average of about 165 cMs per chromosome). The largest wheat map produced to date, by Liu and Tsunewaki (1991), contains about 1800 cMs while the group 3 and group 7 maps produced by Devos et al.

(1992a) and Chao et al. (1989b), respectively, together contain less than 800 cMs.

By using populations segregating for wheat-alien species genetic variation and placing emphasis on the use of gDNA clones (which detect a much higher RFLP level than cDNA clones), low resolution maps of centromeric regions should be fairly readily obtainable. However, obtaining low-resolution maps of the distal regions of chromosomes, where rates of recombination appear to be very high, remains problematic; it may be facilitated by concentrating on specific classes of gDNA clones (Devos et al. 1992a). The evidence obtained to date indicates a very high degree of conservation of locus order among homoeologous wheat chromosomes in centromeric regions and possibly in distal regions as well. This conservation may extend as well to a major degree to other Triticeae genomes, especially in centromeric regions. If this conservation of locus order is confirmed, the necessity of mapping all segments of each individual wheat chromosome will be reduced.

8. References

- Anderson, J A , Oghara, Y , Sorrells, M E and Tanksley, S D (1992) Development of a chromosomal arm map for wheat based on RFLP markers *Theor Appl Genet* 83 1035–1043
- Appels, R and Dvorak, J (1982) The wheat ribosomal DNA spacer region its structure and variation in populations and among species *Theor Appl Genet* 63 337–348
- Appels, R , Gerlach, W R , Dennis, E S , Swift, H and Peacock, W J (1980) Molecular and chromosomal organization of DNA sequences coding for the ribosomal RNA's in cereals *Chromosoma (Berl)* 78 293–311
- Bartels, D , Altosaar, I , Harberd, N P , Barker, R F and Thompson, R D (1986) Molecular analysis of γ -gladin gene families at the complex *Gli-1* locus of bread wheat (*T aestivum* L.) *Theor Appl Genet* 72 845–853
- Baulcombe, D C and Bufford, D (1983) Gibberellic-acid-regulated expression of α -amylase and six other genes in wheat aleurone layers *Planta* 157 493–501
- Baulcombe, D C , Barker, R F and Jarvis, M G (1987a) A gibberellin response wheat gene has homology to yeast carboxypeptidase Y *J Biol Chem* 262 13726–13735
- Baulcombe, D C , Huttly, A K , Martienssen, R A , Barker, R F and Jarvis, M G (1987b) A novel wheat α -amylase gene (*α -Amy3*) *Mol Gen Genet* 209 33–40
- Brinkmann, H , Martinez, D , Quigley, F , Martin, W and Cerff, R (1988) Endosymbiotic origin and codon bias of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize *J Mol Evol* 26 320–328
- Brogie, R , Coruzzi, G , Lamppa, G , Keith, B and Chua, N -H (1983) Structural analysis of nuclear genes coding for the precursor to the small subunit of wheat ribulose-1,5-bisphosphate carboxylase *Bio/Technology* 1 55–61
- Chandler, P M , Zwar, J A , Jacobsen, J V , Higgins, T J V and Inglis, A S (1984) The effects of gibberellic acid and abscisic acid on α -amylase mRNA levels in barley aleurone layers studied using an α -amylase cDNA clone *Plant Mol Biol* 3 407–418
- Chao, S , Sharp, P J and Gale, M D (1988) A linkage map of wheat group 7 chromosomes using RFLP markers In T E Miller and R M D Koebner (eds), *Proc 7th Int Wheat Genet Symp* , pp 493–498 Cambridge, U K
- Chao, S , Rames, C A , Longstaff, M , Sharp, P J , Gale, M D and Dyer, T A (1989a) Chromosomal location and copy number in wheat and some of its close relatives of genes for enzymes involved in photosynthesis *Mol Gen Genet* 218 423–430

- Chao, S., Sharp, P.J., Worland, A.J., Warham, E.J., Koebner, R.M.D. and Gale, M.D. (1989b) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. *Theor. Appl. Genet.* 78: 495–504.
- Cheng, C.L., Dewdney, J., Kleinhofs, J. and Goodman, H.M. (1986) Cloning and nitrate induction of nitrate reductase mRNA. *Proc. Natl. Acad. Sci. U.S.A.* 83: 6825–6828.
- Cheung, W.Y., Moore, G. and Gale, M.D. (1992) *Hpa*II library indicates 'methylation-free islands' in wheat and barley. *Theor. Appl. Genet.* 84: 739–746.
- Chinoy, C.N., Devos, K.M., Bringloe, D., Gray, J.C., Gale, M.D. and Dyer, T.A. (1991) Chromosomal location of the genes for ferredoxin in wheat, barley and rye. *Theor. Appl. Genet.* 82: 1–2.
- Curtis, C.A. and Lukaszewski, A.J. (1991) Genetic linkage between C-bands and storage protein genes in chromosome 1B of tetraploid wheat. *Theor. Appl. Genet.* 81: 245–252.
- Devos, K.M., Atkinson, M.D., Chinoy, C.N., Guiltan, M.J., Quatrano, R.S. and Gale, M.D. (1991a) Chromosomal location and variability in wheat, barley and rye of a wheat gene encoding a β ZIP protein (EmBP-1). *Theor. Appl. Genet.* 82: 665–667.
- Devos, K.M., Chinoy, C.N., Atkinson, M.D., Hansen, L., von Wettstein-Knowles, P. and Gale, M.D. (1991b) Chromosomal location in wheat of the genes coding for the acyl carrier proteins I and III. *Theor. Appl. Genet.* 82: 3–5.
- Devos, K.M., Atkinson, M.D., Chinoy, C.N., Liu, C. and Gale, M.D. (1992a) RFLP-based genetic map of the homoeologous group 3 chromosomes of wheat and rye. *Theor. Appl. Genet.* 83: 931–939.
- Devos, K.M., Atkinson, M.D., Chinoy, C.N., Lloyd, J.C., Raines, C.A., Dyer, T.A. and Gale, M.D. (1992b) The coding sequence for sedoheptulose-1,7-bisphosphatase detects multiple homologues in wheat genomic DNA. *Theor. Appl. Genet.* 85: 133–135.
- Devos, K.M., Millan, T. and Gale, M.D. (1993a) Comparative RFLP maps of the homoeologous group 2 chromosomes of wheat, rye and barley. *Theor. Appl. Genet.* 85: 784–792.
- Devos, K.M. and Gale, M.D. (1993b) Extended genetic maps of the homoeologous group 3 chromosomes of wheat, rye and barley. *Theor. Appl. Genet.* 85: 649–652.
- Doan, N.P. and Fincher, G.B. (1988) The A- and B-chains of carboxypeptidase I from germinated barley originate from a single precursor polypeptide. *J. Biol. Chem.* 263: 11106–11110.
- Dvorak, J. and Appels, R. (1986) Investigation of homoeologous crossing over and sister chromatid exchange in the wheat *Nor-B2* locus coding for rRNA and *Gli-B2* locus coding for gliadins. *Genetics* 113: 1037–1056.
- Dvorak, J. and Chen, K.-C. (1984) Distribution of non-structural variation between wheat cultivars along chromosome arm 6Bp: evidence from the linkage and physical map of the arm. *Genetics* 106: 325–333.
- Dvorak, J., Zhang, H.-B., Kota, R.S. and Lassner, M. (1989) Organization and evolution of the 5S ribosomal RNA gene family in wheat and related species. *Genome* 32: 1003–1016.
- Futers, T.S., Vaughan, T.J., Sharp, P.J. and Cumings, A.C. (1990) Molecular cloning and chromosomal location of genes encoding the 'Early-methionine-labelled' (Em) polypeptide of *Triticum aestivum* L. var. Chinese Spring. *Theor. Appl. Genet.* 80: 43–48.
- Gallii, S., Gallii, G. and Feldman, M. (1991) Chromosomal location of genes for Rubisco small subunit and Rubisco-binding protein in common wheat. *Theor. Appl. Genet.* 81: 98–104.
- Gerlach, W.L. and Bedbrook, J.R. (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucl. Acid Res.* 7: 1869–1886.
- Gerlach, W.L. and Dyer, T.A. (1980) Sequence organization of the repeating units in the nucleus of wheat which contain 5S rRNA genes. *Nucl. Acid Res.* 8: 4851–4865.
- Gill, K.S., Lubbers, E.L., Gill, B.S., Raupp, W.J. and Cox, T.S. (1991) A genetic linkage map of *Triticum tauschii* (DD) and its relationship to the D genome of bread wheat (AABBDD). *Genome* 34: 362–374.
- Guiltan, M.J., Marcotte, W.R. Jr. and Quatrano, R.S. (1990) A plant leucine zipper protein that recognizes an abscisic acid response element. *Science* 250: 267–271.
- Hansen, L. (1987) Three cDNA clones for barley leaf acyl carrier proteins I and III. *Carlsberg Res. Commun.* 52: 381–392.

- Harberd, N P , Bartels, D and Thompson, R D (1985) Analysis of the gliadin multigene loci in bread wheat using nullisomic-tetrasomic lines *Mol Gen Genet* 198 234–242
- Harberd, N P , Bartels, D and Thompson, R D (1986) DNA restriction-fragment variation in the gene family encoding high-molecular-weight (HMW) glutenin subunits of wheat *Biochem Genet* 24 579–596
- Hart, G E (1987) Genetic and biochemical studies of enzymes In E G Heyne (ed), *Wheat and Wheat Improvement*, pp 199–214 American Society of Agronomy, Madison, WI
- Hart, G E and Gale, M D (1988) Guidelines for nomenclature of biochemical/molecular loci in wheat and related species In T E Miller and R M D Koebner (eds), *Proc 7th Int Wheat Genet Symp* , pp 1215–1218 Cambridge, U K
- Huttly, A K , Martienssen, R A and Baulcombe, D C (1988) Sequence heterogeneity and differential expression of the α -Amy2 gene family in wheat *Mol Gen Genet* 214 232–240
- Kam-Morgan, L N W , Gill, B S and Muthukrishnan, S (1989) DNA restriction fragment length polymorphisms a strategy for genetic mapping of D genome of wheat *Genome* 32 724–732
- Kleinhofs, A , Chao, S and Sharp, P J (1988) Mapping of nitrate reductase genes in barley and wheat In T E Miller and R M D Koebner (eds), *Proc 7th Int Wheat Genet Symp* , pp 541–546 Cambridge, U K
- Kreis, M , Williamson, M S , Buxton, B , Pyrell, J , Hejgard, J and Svendsen, I (1987) Primary structure and differential expression of β -amylase in normal and mutant barleys *Eur J Biochem* 169 517–525
- Lagudah, E S , Appels, R , Brown, A H D and McNeil, D (1991) The molecular-genetic analysis of *Triticum tauschii*, the D-genome donor to hexaploid wheat *Genome* 34 375–386
- Lazarus, C M , Baulcombe, D C and Martienssen, R A (1985) Amylase genes of wheat are two multigene families which are differentially expressed *Plant Mol Biol* 5 13–24
- Liu, C J and Gale, M D (1989) *Wsp-1*, a set of genes controlling water-soluble proteins in wheat and related species *Genet Res* 54 173–181
- Liu, C J and Gale, M D (1991) The chromosomal location of genes encoding NADH dehydrogenase isozymes in hexaploid wheat and related species *Genome* 34 44–51
- Liu, C J , Chao, S and Gale, M D (1990) The genetic control of tissue-specific peroxidases *Per-1*, *Per-2*, *Per-3*, *Per-4* and *Per-5* in wheat *Theor Appl Genet* 79 305–313
- Liu, C J , Atkinson, M D , Chinoy, C N , Devos, K M and Gale, M D (1992) Nonhomoeologous translocations between group 4, 5 and 7 chromosomes within wheat and rye *Theor Appl Genet* 83 305–312
- Liu, Y-G and Tsunewaki, K (1991) Restriction fragment length polymorphism (RFLP) analysis in wheat II Linkage maps of the RFLP sites in common wheat *Jpn J Genet* 66 617–633
- Longstaff, M , Raines, C A , McMorro, E M , Bradbeer, J W and Dyer, T A (1989) Wheat phosphoglycerate kinase evidence for recombination between the genes for chloroplastic and cytosolic enzymes *Nucl Acid Res* 17 6569–6580
- Marana, C , Garcia-Olmedo, F and Carbonero, P (1988) Linked sucrose synthase genes in group-7 chromosomes in hexaploid wheat (*Triticum aestivum* L) *Gene* 63 253–260
- May, C E and Appels, R (1987) Variability and genetics of spacer DNA sequences between the ribosomal-RNA genes of hexaploid wheat (*Triticum aestivum*) *Theor Appl Genet* 74 617–624
- McIntosh, R A , Hart, G E and Gale, M D (1989) Catalogue of gene symbols for wheat 1989 supplement *Annu Wheat Newsletter* 35 231–241
- McIntosh, R A , Hart, G E and Gale, M D (1993) Catalogue of gene symbols for wheat In *Proc 8th Int Wheat Genet Symp* (in press)
- Mitchell L E , Dennis, E S and Peacock, W J (1989) Molecular analysis of an alcohol dehydrogenase (Adh) gene from chromosome 1 of wheat *Genome* 32 349–358
- Raikhel, N V and Wilkins, T A (1987) Isolation and characterization of a cDNA clone encoding wheat germ agglutinin *Proc Natl Acad Sci U S A* 84 6745–6749
- Raines, C A , Lloyd, J C , Longstaff, M , Bradley, D and Dyer, T A (1988) Chloroplast fructose-1,6-bisphosphatase the product of a mosaic gene *Nucl Acid Res* 18 7931–7942
- Raines, C A , Longstaff, M , Lloyd, J C and Dyer, T A (1989) Complete coding sequence of wheat

- phosphoribulokinase development and light-dependent expression of the mRNA *Mol Gen Genet* 218 423–430
- Rebmann, G , Hertig, C , Bull, J , Mauch, F and Dudler, R (1991) Cloning and sequencing of cDNA's encoding a pathogen-induced putative peroxidase of wheat (*Triticum aestivum* L) *Plant Mol Biol* 16 329–331
- Rohde, W , Becker, D and Salamini, F (1988) Structural analysis of the waxy locus from *Hordeum vulgare* *Nucl Acid Res* 16 7185–7186
- Sharp, P J and Soltes-Rak, E (1988) Homoeologous relationships between wheat group 2 chromosome arms as determined by RFLP analysis In T E Miller and R M D Koebner (eds), *Proc 7th Int Wheat Genet Symp* , pp 635–637 Cambridge, U K
- Sharp, P J , Desai, S and Gale, M D (1988) Isozyme variation and RFLPs at the β -amylase loci in wheat *Theor Appl Genet* 76 691–699
- Sharp, P J , Chao, S , Desai, S and Gale, M D (1989) The isolation, characterization and application in the Triticeae of a set of wheat RFLP probes identifying each homoeologous chromosome arm *Theor Appl Genet* 78 342–348
- Smith, S M , Bedbrook, J and Speir, J (1983) Characterization of three cDNA clones encoding different mRNAs for the precursor of the small subunit of wheat ribulose bisphosphate carboxylase *Nucl Acid Res* 11 8719–8734
- Snape, J W , Flavell, R B , O'Dell, M , Hughes, W G and Payne, P I (1985) Intra-chromosomal mapping of the nucleolar organiser region relative to three marker loci on chromosome 1B of wheat (*Triticum aestivum*) *Theor Appl Genet* 69 263–270
- Thompson, R D , Bartels, D , Harberd, N P and Flavell, R B (1983) Characterization of the multigene family coding for HMW glutenin subunits in wheat using cDNA clones *Theor Appl Genet* 67 87–96
- Wang, M L , Atkinson, M D , Chinoy, C N , Devos, K M , Harcourt, R L , Liu, C J , Rogers, W J and Gale, M D (1991) RFLP-based genetic map of rye (*Secale cereale* L) chromosome 1R *Theor Appl Genet* 82 174–178
- Wang, M L , Atkinson, M D , Chinoy, C N , Devos, K M and Gale, M D (1992) Comparative RFLP-based genetic maps of barley chromosome 5 (1H) and rye chromosome 1R *Theor Appl Genet* 84 339–344
- Williamson, M S , Ford, J and Kreis, M (1988) Molecular cloning of two isoform inhibitor forms of chymotrypsin inhibitor 1 (CI-1) from barley endosperm and their expression in normal and mutant barleys *Plant Mol Biol* 10 521–535

22. Development of RFLP-based linkage maps in diploid and hexaploid oat (*Avena* sp.)

L.S. O'DONOUGHUE^{1,*}, P.J. RAYAPATI², S.F. KIANIAN³,
M.E. SORRELLS¹, S.D. TANKSLEY¹, M. LEE², H.W. RINES^{3,4}
and R.L. PHILLIPS³

¹ Department of Plant Breeding and Biometry, Cornell University, Ithaca, NY 14853-1901, U S A ,

² Department of Agronomy, Iowa State University, Ames, IA 50011, U S A , ³ Department of Agronomy and Plant Genetics, University of Minnesota, St Paul, MN 55108 U S A ⁴ USDA-ARS, University of Minnesota, St Paul, MN 55108, U S A , * Present address Agriculture Canada, Research Branch, Plant Research Center, Ottawa, Ontario, Canada

Contents

1	Introduction	359	3 1	Aneuploid analysis	365
2	Diploid maps	360	3 2	RFLP map	367
2 1	<i>A atlantica</i> × <i>A hirtula</i> map	360	4	Conclusions	368
2 2	<i>A strigosa</i> × <i>A wiestu</i> map	362	5	Acknowledgments	372
2 3	Comparison of diploid oat maps	364	6	References	372
3	Hexaploid	365			

1. Introduction

The genus *Avena* is organized into 14 taxa, each representing a biological species (8 diploid, 5 tetraploid, and 1 hexaploid taxa) classified on the basis of chromosome number, genome, diaspore (unit of dispersal), flower morphology, and cross fertility (Ladizinsky 1989). Based on chromosome pairing and structure, diploid, tetraploid and hexaploid species were given the genomic designations A or C, AABB or AACC, and AACCCDD, respectively (Rajhathy and Thomas 1974). The primary cultivated oat species are hexaploid ($2n = 6x = 42$) *A. sativa* L. and *A. byzantina* C. Koch. Extensive cytological work has led to the development of karyotypes, as well as aneuploid stocks (Rajhathy and Thomas 1974; Hacker and Riley 1965; Morikawa 1985; Linares et al. 1992; Jellen et al. 1993a, b). However, identification of homoeologous groupings in hexaploid oat has been hindered by the lack of a complete aneuploid series, useful genetic markers, and easily identifiable chromosome morphology. The C-banding technique, and in certain cases the use of semi-automated digital image analysis/enhancement systems (Jellen et al. 1993a), has greatly facilitated the discrimination of individual chromosomes and/or genomes in diploid (Yen and Filion 1977; Fominaya et al. 1988a), tetraploid (Fominaya et al. 1988b) and hexaploid species (Linares et al. 1992; Jellen et al. 1993a). For instance, the C genome chromosomes are easily distinguished on the basis of their dark staining pattern from the A, B or D genome chromosomes (Fominaya et al. 1988; Jellen 1992).

Even though a significant number of physiological and metabolic studies have been conducted using oat, genetic studies have lagged behind those of other plant species (Guerin and Guerin 1993). However, recent advances in oat cytology (Rines et al. 1992), transformation (Somers et al. 1992), isolation and characterization of important genes (Martin et al. 1992), use of maize pollen in the derivation of partially fertile haploids and maize alien chromosome addition lines (Riera-Lizarazu et al. 1993) have made the study and characterization of oat genomes imperative.

The development of DNA markers has allowed the construction of Restriction Fragment Length Polymorphism (RFLP) maps in many crop species. Due to the complexity of genetic analyses and segregation patterns in hexaploids (Sorrells 1992), a number of strategies have been followed to simplify the development of a cultivated oat genomic map. Diploid oat species corresponding to ancestors of hexaploid oat have been used to construct RFLP maps (O'Donoghue et al. 1992; Rayapati et al. 1994). The use of less complex diploid maps to predict linkage arrangements in hexaploids is based on the assumption that the genomes are relatively homosequential. Localization of RFLP markers to aneuploid stocks is another approach taken to reduce the complexities of mapping the large polyploid genome (Jellen 1992; Kianian et al. 1992; Rooney 1992; Wu et al. 1992). In this report, we summarize the mapping efforts of a consortium of scientists at Agriculture Canada-Ottawa, Cornell University, Iowa State University, and the University of Minnesota towards the development of an RFLP map for cultivated oat.

2. Diploid maps

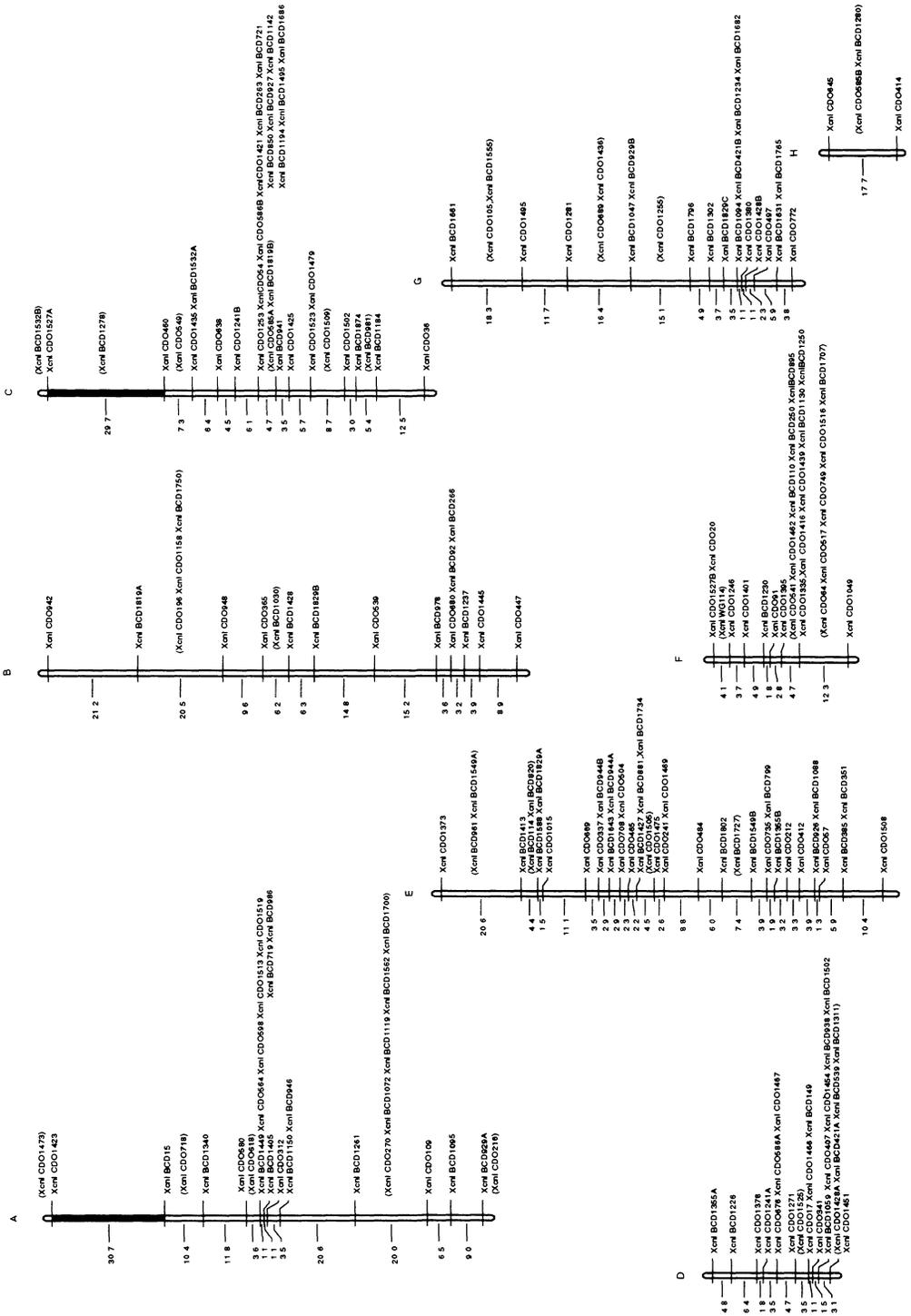
Two independent diploid maps have been constructed. In both cases 'A-genome' diploid species were used. The first map was based on 44 F_3 families from a cross of *A. atlantica* Baum et Fedak \times *A. hirtula* Lag. (O'Donoghue et al. 1992) and the other, on an F_2 population (88 individuals) of *A. strigosa* Schreb. \times *A. wiestii* Steud. (Rayapati et al. 1994). These maps are being integrated to develop a composite RFLP linkage map of diploid oat.

2.1. *A. atlantica* \times *A. hirtula* map

A total of 194 RFLP markers were used in the construction of the *A. atlantica* \times *A. hirtula* map. The main source of markers were oat (CDO) and barley (BCD) leaf cDNA libraries. Of the 194 RFLP loci, only two remained unlinked and seven main linkage groups, presumably corresponding to the seven

→

Fig. 1. 'A-genome' oat RFLP map based on an F_2 of the cross *A. atlantica* \times *A. hirtula* (O'Donoghue et al. 1992). Map distances are given in centi-Morgans [Kosambi function (Kosambi 1944)]. Linkages between markers of > 25 cM are in solid black. Markers in parentheses have been assigned to intervals only.



chromosomes of the haploid genome, were identified (Fig. 1). A small linkage group (group H) formed by four clones which showed loose linkage to both group A and group D was also identified (Fig. 1). The actual and estimated map sizes obtained were 614 and 764 cM, respectively. These values are somewhat smaller than the map sizes obtained in other species of similar genome size and the *A. strigosa* × *A. wiestii* diploid oat map. Possibly, the small genome size can be attributed to reduced recombination associated with the interspecific nature of the cross. In addition to this apparent general reduction in recombination, specific areas of tightly clustered markers were also noted. Linkage groups A, C, D, F and G exhibited distinct areas with several markers that were closely spaced or co-segregating (Fig. 1). These areas may represent reduced recombination around the centromeres. Reduced recombination around centromeres is known to occur in several species such as maize (Rhoades 1955) and tomato (Tanksley et al. 1992) and may be due to the centromere itself or to centric heterochromatin (Roberts 1965). C-banding of 'A-genome' *Avena* species shows no evidence of centromeric heterochromatin in mitotic chromosomes (Fominaya et al. 1988; Jellen 1992), although large heterochromatic blocks are seen around the centromeres of all chromosomes at pachytene in hexaploid oat (Johnson et al. 1987).

Thirteen of the mapped clones identified duplicated sequences within this diploid genome; half of these had one of their loci on linkage group D and another on either C, G or E. In addition to single locus duplication, segmental duplications were also evident. Nineteen percent of the markers showed significant deviation from the expected Mendelian ratios. Many of these markers 18/37 (49%) were located on linkage group A, where 64% of all markers (11/18) gave skewed inheritance ratios. Except for one, all of these markers were skewed toward the male parent *A. hirtula*. Similar trends for clustering and preferential transmission were observed in other parts of the genome. These distortions could be due to genes affecting gamete or hybrid viability and/or inadvertent selection due to the small size of the population.

2.2. *A. strigosa* × *A. wiestii* map

An F₂ population was produced by crossing the diploid species *A. strigosa* (CI 3815) with *A. wiestii* (CI 1994) which are resistant and susceptible, respectively, to 40 isolates of *Puccinia coronata*, the causal agent of crown rust. Eighty-eight F₂ individuals were used to construct a RFLP linkage map. Two hundred and ten RFLP loci have been placed in eleven linkage groups (Fig. 2). One-hundred and seventy clones that detect 178 loci were from an oat root cDNA library (ISU) and 30 were selected from other oat marker sets. This map covers 2470 cM, with an average of 12 cM between RFLP loci (Fig. 2).

Eighty-eight F₃ lines, derived from the F₂ individuals used to construct the map, were screened for resistance to nine *P. coronata* isolates. The resistant phenotype in response to eight isolates is dominant and each is determined by alleles at a single locus designated *Pca*. Resistance to a ninth isolate was

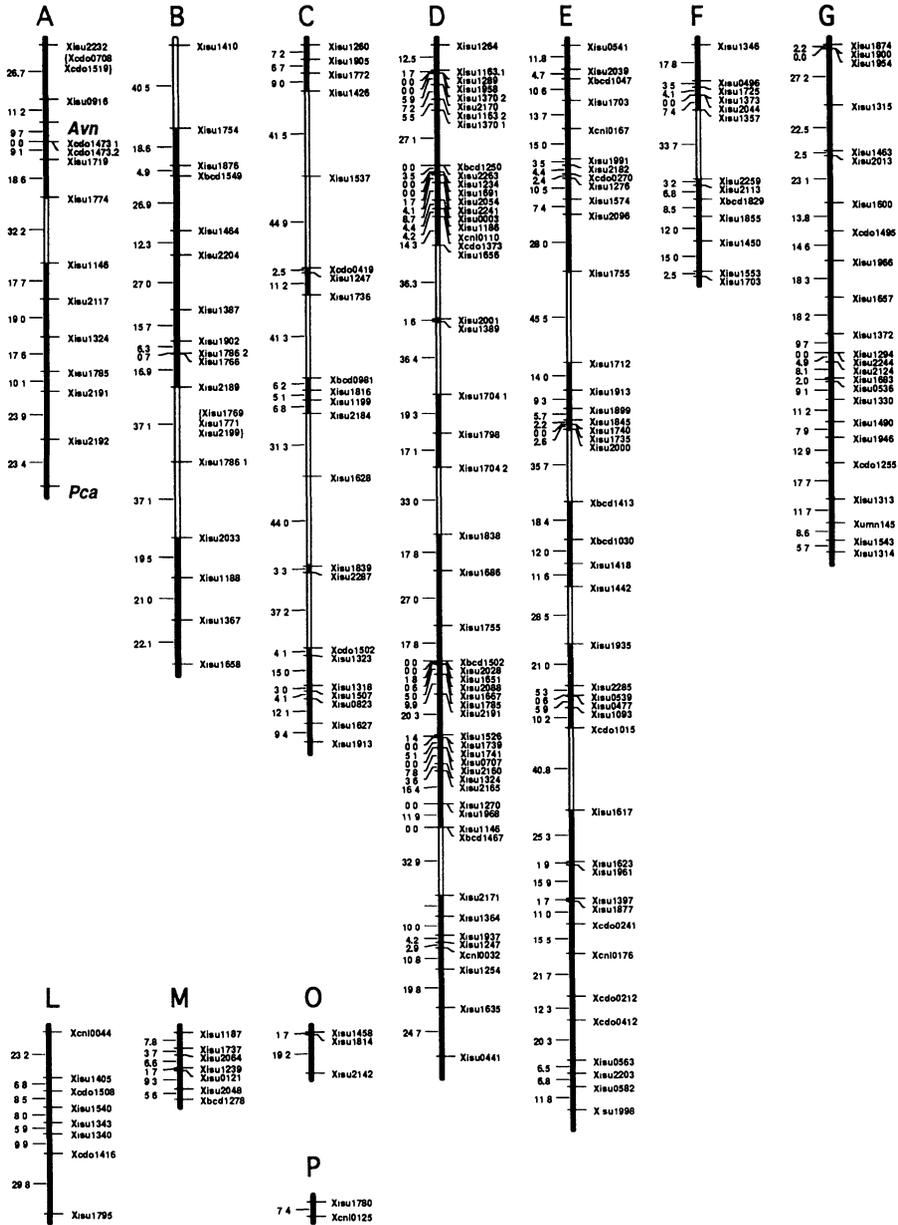


Fig 2 Genetic linkage map of diploid *Avena* representing the A genome. The 212 RFLP loci, one avenin locus (*Avn*), and one *Pca* locus conferring resistance to at least 9 isolates of *Puccinia coronata* were assigned to 11 linkage groups (A-G, L, M, O, and P). Numbers to the left of a linkage group represent map distances in cM calculated with the Kosambi function. Designations to the right are locus names. Names beginning with an X are RFLP loci. Open bars represent regions between groups of markers linked by a LOD score greater than or equal to 3.0, but separated by more than 30 cM. Solid bars represent regions between markers linked by less than 30 cM at LOD 5.0.

probably regulated by two dominant epistatic genes, one of which maps to the *Pca* locus. Experiment with recombinant inbred lines (RILs) demonstrate that the *Pca* locus confers resistance to 14 isolates of *P. coronata*. Recombination within the *Pca* locus have been detected using RILs demonstrating that it is a complex locus of at least three genes (Rayapati et al. unpub.). Genetic stocks identifying each resistance gene can be prepared from F₇ RILs of this mapping population. The RFLP markers and genetic stocks generated by this study provide valuable tools for further investigations of the genetics of host-pathogen interactions in *Avena* and for related grasses such as *Triticum*, *Hordeum*, and *Secale*.

2.3. Comparison of diploid oat maps

One of the major differences noted between the two diploid oat maps is the genome size. Greater recombination frequencies and, therefore, greater map distances, have been observed for intraspecific crosses than for interspecific crosses of rice (McCouch et al. 1988), potato (Gebhardt et al. 1991), tomato (Rick 1969), maize (Doebley and Stec 1991), and *Brassica* (Kianian and Quiros 1992). When interspecific *Lycopersicon esculentum* × *L. pennellii* crosses were compared to intraspecific *L. esculentum* × *L. esculentum* controls, recombination rates varied, according to chromosomal segments, from 10% to 100% of the intraspecific control values (Rick 1969). Mapping studies comparing genetically divergent interspecific crosses of teosinte and maize to intraspecific crosses between inbred lines of maize show that maps generated by the intraspecific populations were two to five times larger than the interspecific ones (Doebley and Stec 1991). Chromosome aberrations can explain reduced recombination frequencies in interspecific crosses (Kianian and Quiros 1992; Vallejos et al. 1992); however, reduced recombination was observed in interspecific crosses even when detailed cytogenetic analysis revealed normal pairing and chiasmata formation (Rick 1969). These observations suggest that the genomes of intraspecific crosses undergo greater recombination than those of interspecific genomes.

The frequency of RFLPs between *A. atlantica* and *A. hirtula* detected with single digests of two enzymes (*Eco*RI and *Eco*RV) was 74% (O'Donoghue et al. 1992), while only 39% of clones detected RFLPs in the *A. strigosa* and *A. wiestii* parents under the same conditions. Thus, *A. strigosa* and *A. wiestii* are possibly less divergent than *A. atlantica* and *A. hirtula* and likely represent the same species. This is consistent with Ladizinsky's (1989) estimate of the evolutionary divergence of these taxa describing *A. strigosa*, *A. wiestii*, and *A. hirtula* as morphological variants belonging to one biological species *A. strigosa*, while *A. atlantica* is a member of a distinct biological species.

To assess the possibility of different recombination frequencies in different diploid oat crosses, 26 probes that detected RFLPs in the *A. atlantica*/*A. hirtula* cross were mapped in the *A. strigosa*/*A. wiestii* cross. Thirteen of these probes showed conserved synteny in both populations (Fig. 2, Table 1). The 3.5 fold

Table 1 Recombination among syntenic markers in two diploid oat maps^a

Syntenic markers	<i>A. strigosa/A. wiestii</i>	<i>A. atlantica/A. hirtula</i>
CDO1467-BCD1502	78cM	10.8cM
CDO1473-CDO1519	70cM	56.5cM
CDO1495-CDO1255	117cM	43.2cM
CDO1502-BCD981	134cM	8.4cM
BCD1413-CDO1015	113cM	5.9cM
CDO1015-CDO241	97cM	32.0cM
CDO241-CDO212	37cM	31.2cM
CDO212-CDO412	12cM	3.3cM
Total	658cM	191 cM
Total map length	2470cM	614 cM

^a Map distances are given in centi-Morgan calculated by the Kosambi function (Kosambi 1944)

reduction in recombination for regions detected by these probes in the *A. atlantica/A. hirtula* cross is consistent with the four fold decrease (2470 vs. 614 cM) in total length for the *A. atlantica/A. hirtula* map (Table 1). The difference in map sizes produced by two different diploid oat crosses is consistent with previous comparisons of maps produced from intra- and interspecific crosses in maize and wheat. Doebley and Stec (1991) identified eight pairs of markers that were linked by less than 25 cM on an interspecific maize/teosinte map. The same pair of loci were linked by more than 40 cM on an intraspecific map. Without intervening markers, these pairs could have been considered to be unlinked in the intraspecific maize map. The diploid oat map (Fig. 2) depicts 20 regions linked by more than 30 cM (unshaded regions). These linkages have been tentatively assigned because the adjacent markers are at least 1000 times more likely to be linked to each other (LOD = 3.0) than to any other markers on this map. Two maps of diploid wheat representing the D genome show gaps greater than 30 cM but were shown to be physically linked using aneuploid stocks (Gill et al. 1991; Lagudah et al. 1991). The likelihood that the *A. strigosa/A. wiestii* map is constructed from an intraspecific cross while the *A. atlantica/A. hirtula* map is constructed from a much more divergent interspecific cross is the best explanation for the greater recombination frequency, larger gaps, and more extensive genome coverage of the *A. strigosa/A. wiestii* map.

3. Hexaploid

3.1. Aneuploid analysis

Aneuploid stocks have been used to localize RFLP markers to a specific chromosome or chromosome arm in several crops including wheat (Devey and Hart 1988; Lagudah et al. 1991; Gill et al. 1991; Anderson et al. 1992), rice

HOMOELOGOUS GROUP ASSIGNMENT

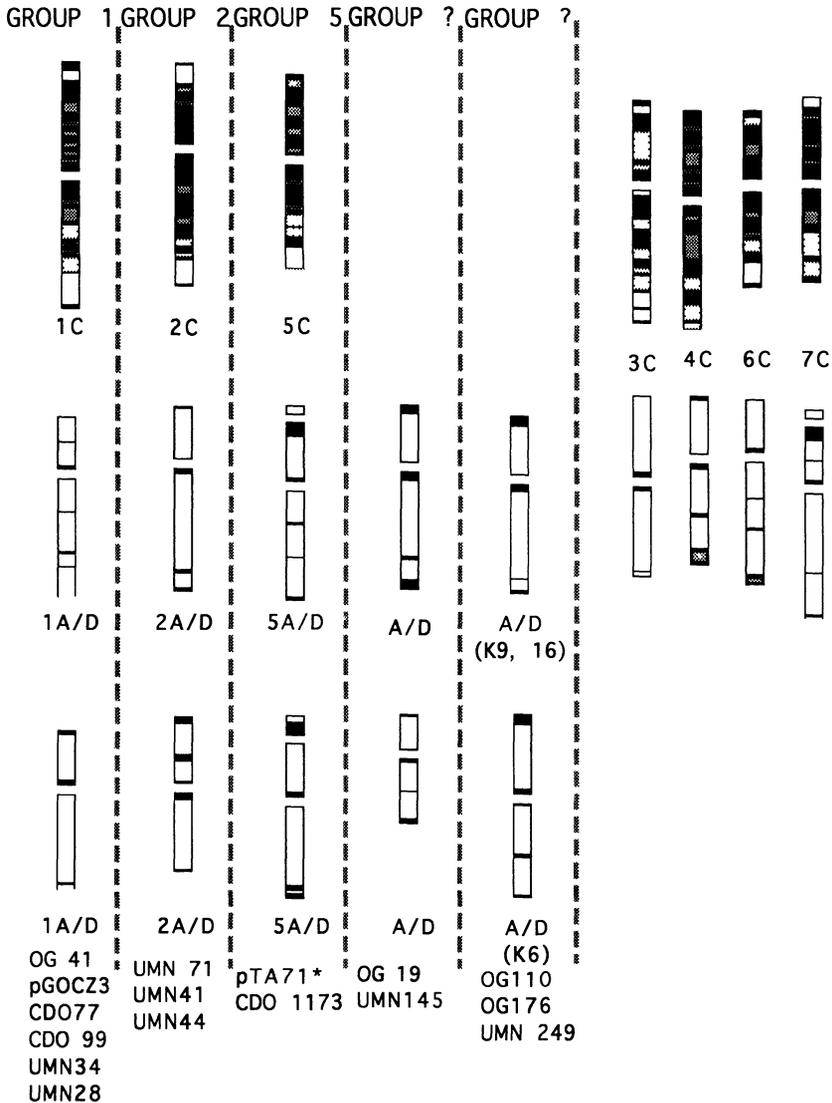


Fig 3 Homoeologous chromosome sets of hexaploid oat identified by C-banding pattern and RFLP analysis of aneuploid stocks. The C-banding ideogram (Jellen 1992) and RFLP markers assigned to each group are depicted.

(McCouch et al. 1988) and maize (Helentjaris et al. 1986). In hexaploid cultivated oat, aneuploids including nullisomics ($2n-2 = 40$), monosomics ($2n-1 = 41$) and ditelosomics ($2n-tt = 40 + tt$) have been recovered. Nullisomics have been effective for assigning RFLP markers to specific chromosomes (Kianian et al. 1992; Mendu et al. 1993) and ditelosomics allow

for localization of such markers to chromosome arms (Wu et al. 1992). Since a complete set of oat nullisomic stocks is not available, monosomics may be used to map markers to chromosomes by dosage analysis (Rooney 1992). A DNA sequence on a monosomic chromosome should produce a Southern blot band with reduced intensity when compared to the band from a line disomic for the same chromosome. Nullisomic and ditelosomic lines of *A. sativa* cv SunII and monosomic lines of *A. byzantina* cv Kanota have been used to assign 129 RFLP markers to sixteen syntenic groups (Kianian et al. 1992; Rooney 1992; Wu et al. 1992). Determination of map position for some of these sequences has allowed tentative localization of 15 linkage groups to ten physical chromosomes. In the simplest case, sequences complementary to a single copy RFLP probe should produce from one to three bands, mapping to a homoeologous chromosome set, in an allohexaploid such as oat. Aneuploid analysis with RFLPs in conjunction with the C-banding technique has allowed the tentative identification of 3 complete and 2 partial (2/3) sets of homoeologous chromosomes (Fig. 3).

3.2. *RFLP map*

A recombinant inbred F_6 -derived population of 71 individuals from a cross of *A. byzantina* cv Kanota \times *A. sativa* cv Ogle is being used in the development of the cultivated oat map (O'Donoghue et al., unpub.). To date over 560 RFLP sequences derived mostly from four cDNA libraries constructed from etiolated leaf tissue of oat (CDO) and barley (BCD), developing oat endosperm (UMN) and oat root tissue (ISU) have been detected on the mapping population. Thirty-three linkage groups of 3 or more markers, five groups with 2 markers each and 29 unlinked markers have been identified (Fig. 4). Although incomplete, this map reveals several interesting features about the oat genome.

This difference may indicate that oat genome contains more duplicated sequences. The current assignment of at least thirty-eight linkage groups, 17 more than expected ($n = 3x = 21$), despite the large number of markers already on the map, may indicate large heterochromatic regions that are recombinogenic but lack active genes. Such areas presumably do not contribute to the cDNA libraries used in the map construction and are, therefore, not marked. Further evidence to support this hypothesis is that recently added cDNA probes tend to map within or extend identified linkage groups rather than to connect them, aneuploid analysis places 15 linkage groups to ten physical chromosomes (Kianian et al. 1992) and the placement of centromere distal in a linkage group with no obvious clustering of nearby markers (Wu et al. 1992). Therefore, strategies are being devised for connecting these linkage groups such as the use of genomic or chromosome-specific libraries, bulked segregant analysis with random primers, aneuploid assignment using newly derived stocks from oat haploids (Davis 1992), and populations derived from interspecific crosses. Distinct areas with several markers closely spaced or co-segregating can be seen in some of the linkage groups (Fig. 4). These areas of reduced recombination could represent functionally active regions (i.e. many

expressed genes) which are saturated by the four cDNA libraries used, cytological aberrations (such as heterozygous inversions and translocations), or simply areas with naturally low recombination rates such as regions near centromeres. Numerous cytological data (McMullen et al. 1982; Singh and Kolb 1991; Kianian et al. 1992) have indicated a common occurrence of reciprocal translocation(s) between various oat cultivars. Meiotic analyses of Kanota \times Ogle F₁ hybrids have also indicated the existence of at least one reciprocal translocation (i.e. ring of four chromosomes in diakinesis). The meiotic analysis and the disappearance of parental alleles in some of the F₆ lines (i.e. duplicate-deficient [Dp-Df] lines) suggest that the region of reduced recombination on at least one of the linkage groups is due to translocation heterozygosity (Fig. 4).

One hundred and two of the clones used on the hexaploid oat map had previously been mapped on the *A. atlantica* \times *A. hirtula* diploid oat map (O'Donoghue et al. 1992). Comparisons of marker order between the two maps reveal that although large groups of markers remain syntenic, at least 5 major translocation differences exist between the A genome diploids and the hexaploid oat cultivars mapped. It is apparent that within the genus *Avena* many structural rearrangements have taken place during the course of evolution. This fact may explain past difficulties in obtaining a complete aneuploid series in oat as well as in the clear characterization of individual genomes.

4. Conclusions

Construction of linkage maps in polyploids has lagged behind that of diploid species due to the complexities of genome size and genetic analysis (Sorrells 1992). The development of diploid oat maps and chromosome arm maps based on aneuploids concurrently with the development of a hexaploid oat map is providing a more comprehensive picture of genome organization in the genus *Avena*. This information is essential as it is becoming increasingly clear that *Avena* is characterized by frequent structural rearrangements both between and within species. However, these structural differences may offer oat geneticists certain advantages. Translocations identified cytologically are being mapped genetically with RFLP markers and these markers can be used to identify duplication-deficiency (Dp-Df) lines. These lines can be employed in the generation of specific gene duplications such as those for crown rust resistance

→

Fig. 4. The hexaploid oat RFLP map based on the cross *A. byzantina* cv. Kanota \times *A. sativa* cv. Ogle (O'Donoghue et al., in preparation). Map distances are given in centi-Morgans [Kosambi function (Kosambi 1944)]. Markers in parentheses have been assigned to their most likely interval location as precise placement did not meet the LOD threshold of 2.0.

(Wilson and McMullen 1992). The difference in recombination frequencies (about 4 fold) between the two diploid populations raises interesting questions about chromosome pairing, crossing over and recombination frequencies in specific crosses and factors affecting these properties. The relationship between chromosome pairing in inter- and intraspecific crosses and recombination frequency is currently being investigated. The cultivated oat mapping project will provide a set of recombinant inbred lines which are well characterized both genetically and cytologically. This material will be invaluable for further genetic studies in oat such as the identification of chromosomal regions controlling traits of agronomic and biological significance.

5. Acknowledgments

The authors wish to thank the scientists who provided their unpublished results for inclusion in this chapter. We are also grateful to a number of technical staff at the four institutions for providing expertise throughout this project. The support of The Quaker Oats Company is especially appreciated.

6. References

- Anderson, J A , Ogihara, Y , Sorrells, M E and Tanksley, S D (1992) Development of a chromosomal arm map for wheat based on RFLP markers *Theor Appl Genet* 83 1035–1043
- Davis, D W. (1992) Characterization of oat haploids and their progeny M S thesis, University of Minnesota, St Paul, MN
- Devey, M and Hart, G (1988) Intergenomic restriction fragment length polymorphism in hexaploid wheat *Agron Abstr* , p 79
- Doebley, J and Stec, A (1991) Genetic analysis of the morphological differences between maize and teosinte *Genetics* 129 285–295
- Fominaya, A , Vega, C and Ferrer, E (1988a) Giemsa C-banded karyotypes of *Avena* species *Genome* 30 627–632
- Fominaya, A , Vega, C and Ferrer, E (1988b) C-banding and nucleolar activity of tetraploid *Avena* species *Genome* 30 633–638
- Gebhardt, C , Ritter, E , Barone, A , Debener, T , Walkemeier, B , Canal, U M W , Tanksley, S D and Salamini, F (1991) RFLP maps of potato and their alignment with the homoeologous tomato genome *Theor Appl Genet* 83 49–57
- Goffreda, J C , Brunquist, W B , Beer, S C , Tanksley, S D and Sorrells, M E (1992) Application of molecular markers to assess genetic relationships among accessions of wild oat, *Avena sterilis* *Theor Appl Genet* 85 146–151
- Gill, K S , Lubbers, E L , Gill, B S , Raupp, W J and Cox, T S (1991) A genetic linkage map of *Triticum tauschii* (DD) and its relationship to the D genome of bread wheat (AABBDD) *Genome* 34 362–374
- Guerin, T F and Guerin, P M (1993) Recent developments in oat molecular biology *Plant Mol Biol Rep* 11 65–72
- Hacker, J B and Riley, R (1965) Morphological and cytological effects of chromosome deficiency in *Avena sativa* *Can J Genet Cytol* 7 304–315
- Helentjaris, T , Weber, D and Wright, S (1986) Use of monosomics to map cloned DNA fragments in maize *Proc Natl Acad Sci U S A* 83 6035–6039

- Jellen, E N (1992) Characterization of image-enhanced, C-banded oat (*Avena* spp) monosomics and identification of oat genomes and homoeologous chromosomes Ph D thesis, University of Minnesota, St Paul, MN
- Jellen, E N, Phillips, R L and Rines, H W (1993a) C-banded karyotypes and polymorphisms in hexaploid oat accessions (*Avena* spp) using Wright's stain *Genome* 36 1129-1137
- Jellen, E N, Rooney, W L, Phillips, R L and Rines, H W (1993b) Characterization of the hexaploid oat *Avena byzantina* cv Kanota monosomic series using C-banding and RFLPs *Genome* 36 962-970
- Johnson, S S, Phillips, R L and Rines, H W (1987) Possible role of heterochromatin in chromosome breakage induced by tissue culture in oats (*Avena sativa* L) *Genome* 29 439-446
- Kianian S F, Phillips, R L and Rines, H W (1992) Genetic recombination estimates in intra- and interspecific crosses of hexaploid oat using RFLP markers *Agron Abstr*, p 192
- Kianian, S F and Quiros, C F (1992) Generation of a *Brassica oleracea* composite RFLP map linkage arrangements among various populations and evolutionary implications *Theor Appl Genet* 84 544-554
- Kosambi, D D (1944) The estimation of map distance from recombination values *Ann Eugen* 12 172-175
- Ladizinsky, G (1989) Biological species and wild genetic resources in *Avena* IBPGR Report of a Working Group on *Avena*, pp 19-32
- Lagudah, E S, Appels, R, Brown, A H D and McNell, D (1991) The molecular genetic analysis of *Triticum tauschii*, the D-genome donor to hexaploid wheat *Genome* 34 375-386
- Linares, C, Vega, C, Ferrer, E and Fominaya, A (1992) Identification of C-banded chromosomes in meiosis and the analysis of nucleolar activity in *Avena byzantina* C Koch cv 'Kanota' *Theor Appl Genet* 83 650-654
- Martin, D J, Somers, D A, Gengenbach, B G and Rines, H W (1992) Expression of oat (1-3, 1-4)- β -glucanase during kernel development *Agron Abstr*, p 193
- McCouch, S, Kochert, G, Yu, Z, Wang, Z, Khush, G, Coffman, W and Tanksley S (1988) Molecular mapping of rice chromosomes *Theor Appl Genet* 76 815-829
- McMullen, M S, Phillips, R L and Stuthman, D D (1982) Meiotic irregularities in *Avena sativa* L/A *sterilis* L hybrids and breeding implications *Crop Sci* 22 890-897
- Mendu, N, Rines, H and Silflow, C D (1993) Mapping of beta-tubulin genomic sequences in hexaploid oat (*Avena sativa* L) *Theor Appl Genet* 86 135-140
- Morikawa, T (1985) Identification of the 21 monosomic lines in *Avena byzantina* C Koch cv 'Kanota' *Theor Appl Genet* 70 271-278
- O'Donoghue, L S, Wang, Z, Roder, M, Kneen, B, Leggett, M Sorrells M E and Tanksley S D (1992) An RFLP-based linkage map of oats based on a cross between two diploid taxa (*Avena atlantica* \times *A hirtula*) *Genome* 35 765-771
- Rajhathy, T and Thomas, H (1974) Cytogenetics of oats (*Avena* L) In A Wilkes (ed), *Miscellaneous Publication of The Genetic Society of Canada-No 2*, Ottawa, Ontario, Canada
- Rayapati, P J, Gregory, J W, Lee, M and Wise, R P (1994) A linkage map of diploid *Avena* based on RFLP loci and locus conferring resistance to nine isolates of *Puccinia coronata* var *avenae* *Theor Appl Genet* (in press)
- Rhoades, M M (1955) The cytogenetics of maize In G F Sprague (ed), *Corn and Corn Improvement*, pp 123-219 Academic Press, New York
- Rick, C M (1969) Controlled introgression of chromosomes of *Solanum pennellii* into *Lycopersicon esculentum* segregation and recombination *Genetics* 62 753-768
- Riera-Lizarazu, O, Rines, H W, Phillips, R L and Davis, D W (1993) Oat \times maize crosses generate oat haploids that retain some maize chromosomes *Abstr 35th Annual Maize Genet Conf*, p 20
- Rines, H, Phillips, R, Jellen, E, Rooney, W, Kianian, S and Wu, B-C (1992) Toward an integrated chromosomal/molecular genetic map in oat *Proc Fourth Int Oat Conf II* 97-100
- Robert, P A (1965) Difference in behaviour of eu- and hetero-chromatin crossing over *Nature* 205 725-726
- Rooney, W L (1992) Identification and characterization of RFLP markers linked to crown rust

- resistance in oat (*Avena* sp) Ph D thesis, University of Minnesota, St Paul, MN
- Singh, R J and Kolb, F L (1991) Chromosomal interchanges in six hexaploid oat genotypes *Crop Sci* 31 726-729
- Somers, D A , Rines, H W , Gu, W , Kaeppler, H F and Bushnell, W R (1992) Fertile, transgenic oat plants *Bio/Technology* 10 1589-1594
- Sorrells, M E (1992) Development and application of RFLPs in polyploids *Crop Sci* 32 1086-1091
- Tanksley, S D , Ganai, M W , Prince, J P , DeVincente, M C , Bonierbale, M W , Broun, P , Fulton, T M , Giovanonni, J J , Grandillo, S , Martin, G B , Messeguer, R , Miller, J C , Miller, L , Paterson, A H , Pineda, O , Roder, M , Wing, R A , Wu, W and Young, N D (1992) High density molecular linkage maps of the tomato and potato genomes biological inferences and practical applications *Genetics* 132 1141-1160
- Vallejos, C E , Sakiyama, N S and Chase, C D (1992) A molecular marker-based linkage map of *Phaseolus vulgaris* L *Genetics* 131 733-740
- Wilson, W and McMullen, M (1992) Use of a chromosome translocation for genetic studies in oat *Proc Fourth Int Oat Conf I* 146-148
- Wu, B C , Kianian, S F , Phillips, R L and Rines, H W (1992) Mapping RFLP markers relative to the centromere by ditelosomic analysis in oat *Agron Abstr* , p 199
- Yen, S-T and Filon, W G (1977) Differential Giemsa staining in plants V Two types of constitutive heterochromatin in species of *Avena* *Can J Genet Cytol* 19 739-743

List of contributors

JEFFREY L. BENNETZEN, Department of Biological Sciences, Purdue University, W. Lafayette, IN 47907, U.S.A.

J.H. BOUTON, Department of Agronomy, University of Georgia, Athens, GA 30602, U.S.A.

E.C. BRUMMER, Department of Agronomy, University of Georgia, Athens, GA 30602, U.S.A. (Present address: Department of Agronomy, Iowa State University, Ames, IA 50011, U.S.A.)

BENJAMIN BURR, Biology Department, Brookhaven National Laboratory, Upton, NY 11973, U.S.A.

SAM CARTINHOOR, Departments of Genetics, Harvard Medical School and Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, U.S.A.

J. MICHAEL CHERRY, Departments of Genetics, Harvard Medical School and Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, U.S.A.

EDWARD H. COE, Plant Genetics Research Unit, ARS-USDA, and Department of Agronomy, University of Missouri, Columbia, MO 65211, U.S.A.

G. CSANADI, Institute of Genetics, Biological Research Center, Hungarian Academy of Sciences, Szeged, P.O. Box 521, H-6701, Hungary

JOHN DOEBLEY, Department of Plant Biology, University of Minnesota, St. Paul, MN 55108, U.S.A.

C.S. ECHT, Department of Plant and Soil Science, Montana State University, Bozeman, MT 59717-0312, U.S.A.

JACK M. GARDINER, Plant Genetics Research Unit, ARS-USDA, and Department of Agronomy, University of Missouri, Columbia, MO 65211, U.S.A.

CHRISTIANE GEBHARDT, Max-Planck-Institut für Züchtungsforschung, Carl-von-Linné-Weg, D-50829, Köln, Germany.

HOWARD M. GOODMAN, Departments of Genetics, Harvard Medical School and Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, U.S.A.

J. GYÖRGYÉY, Institute of Genetics, Biological Research Center, Hungarian Academy of Sciences, Szeged, P.O. Box 521, H-6701, Hungary

TRACY HALWARD, Department of Botany, University of Georgia, Athens, GA 30602-7271, U.S.A.

SUSAN HANLEY, Departments of Genetics, Harvard Medical School and Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, U.S.A.

GARY E. HART, Soil and Crop Sciences Department, Texas A&M University, College Station, TX 77843, U.S.A.

BRIAN HAUGE, Departments of Genetics, Harvard Medical School and Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, U.S.A.

JINGUO HU, Department of Vegetable Crops, University of California, Davis, CA 95616, U.S.A.

K. KALMAN, Institute of Genetics, Biological Research Center, Hungarian Academy of Sciences, Szeged, P.O. Box 521, H-6701, Hungary

SHERRY KEMPIN, Division of Biology, 156-29, California Institute of Technology, Pasadena, CA 91125, U.S.A.

RICHARD V. KESSELI, Department of Biology, University of Massachusetts, Harbor Campus, Boston, MA 02125-3393, U.S.A.

S.F. KIANIAN, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, U.S.A.

K.K. KIDWELL, Department of Agronomy, University of Wisconsin, Madison, WI 53706, U.S.A.

ANDRZEJ KILIAN, Departments of Crop & Soil Sciences and Genetics & Cell Biology, Washington State University, Pullman, WA 99164–6420, U.S.A. and Department of Crop & Soil Sciences and Department of Genetics, Silesian University, 40-032, Katowice, Poland.

G.B. KISS, Institute of Genetics, Biological Research Center, Hungarian Academy of Sciences, Szeged, P.O. Box 521, H-6701, Hungary

ANDRIS KLEINHOF, Departments of Crop & Soil Sciences and Genetics & Cell Biology, Washington State University, Pullman, WA 99164–6420, U.S.A.

STEVEN J. KNAPP, Department of Crop and Soil Science and Center for Gene Research and Biotechnology, Oregon State University, Corvallis, OR 97331, U.S.A.

GARY KOCHERT, Department of Botany, University of Georgia, Athens, GA 30602, U.S.A.

MAARTEN KOORNNEEF, Department of Genetics, Agricultural University of Wageningen, Dreijenlaan 2, 6703 HA Wageningen, The Netherlands.

M. LEE, Department of Agronomy, Iowa State University, Ames, IA 50011, U.S.A.

T.J. McCOY, Department of Plant and Soil Science, Montana State University, Bozeman, MT 59717-0312, U.S.A.

LEONARD MEDRANO, Division of Biology, 156-29, California Institute of Technology, Pasadena, CA 91125, U.S.A.

ADMASU MELAKE-BERHAN, Department of Biological Sciences, Purdue University, W. Lafayette, IN 47907, U.S.A.

ELLIOT MEYEROWITZ, Division of Biology, 156-29, California Institute of Technology, Pasadena, CA 91125, U.S.A.

RICHARD W. MICHELMORE, Department of Vegetable Crops, University of California, Davis, CA 95616, U.S.A.

L.S. O'DONOUGHUE, Department of Plant Breeding and Biometry, Cornell University, Ithaca, NY 14853-1901, U.S.A. (Present address: Agriculture Canada, Research Branch, Plant Research Center, Ottawa, Ontario, Canada.)

L. ÖKRÉSZ, Institute of Genetics, Biological Research Center, Hungarian Academy of Sciences, Szeged, P.O. Box 521, H-6701, Hungary

T.C. OSBORN, Department of Agronomy, University of Wisconsin, Madison, WI 53706, U.S.A.

RONALD L. PHILLIPS, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108-6026, U.S.A.

CARLOS F. QUIROS, Department of Vegetable Crops, University of California, Davis, CA 95616, U.S.A.

A.E. RACZKEVY, Institute of Genetics, Biological Research Center, Hungarian Academy of Sciences, Szeged, P.O. Box 521, H-6701, Hungary

P.J. RAYAPATI, Department of Agronomy, Iowa State University, Ames, IA 50011, U.S.A.

H.W. RINES, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, U.S.A. and USDA-ARS, University of Minnesota, St. Paul, MN 55108, U.S.A.

ENRIQUE RITTER, Granja Modelo – CIMA, km 356 de la N.I., Arkaute (Alava), Apartado Correas 46, E-01080 Vitoria-Gasteiz, Spain.

EDWARD J. RYDER, USDA, ARS, U.S. Agricultural Research Station, 1636 East Alisal, Salinas, CA 93905, U.S.A.

FRANCESCO SALAMINI, Max-Planck-Institut für Züchtungsforschung, Carl-von-Linné-Weg, D-50829, Köln, Germany

RANDY C. SHOEMAKER, USDA-ARS-FCR, Department of Agronomy, Iowa State University, Ames, IA 50011, U.S.A.

M.E. SORRELLS, Department of Plant Breeding and Biometry, Cornell University, Ithaca, NY 14853-1901, U.S.A.

H. THOMAS STALKER, Department of Crop Science, North Carolina State University, Raleigh, NC 27695-3281, U.S.A.

PIET STAMM, Department of Genetics, Agricultural University of Wageningen, Dreijenlaan 2, 6703 HA Wageningen, The Netherlands.

CHARLES W. STUBER, U.S. Department of Agriculture, Agricultural Research Service, Department of Genetics, North Carolina State University, Raleigh, NC 27695-7614, U.S.A.

S.D. TANKSLEY, Department of Plant Breeding and Biometry, Cornell University, Ithaca, NY 14853-1901, U.S.A.

MARIA JOSE TRUCO, Department of Vegetable Crops, University of California, Davis, CA 95616, U.S.A. (Present address: IRTA, Corta de Cabrils S/N, 08348 Cabrils, Spain.)

C. EDUARDO VALLEJOS, Department of Horticultural Sciences & Graduate Program in Plant Molecular and Cellular Biology, 1255 Fifield Hall, University of Florida, Gainesville, FL 32611, U.S.A.

INDRA K. VASIL, Laboratory of Plant Cell and Molecular Biology, University of Florida, Gainesville, FL 32611-0690, U.S.A.

JONATHAN F. WENDEL, Department of Botany, Iowa State University, Ames, IA 50011, U.S.A.

RICHARD WHITKUS, Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, U.S.A.

NEVIN DALE YOUNG, Department of Plant Pathology, 495 Borlaug Hall, University of Minnesota, St. Paul, MN 55108, U.S.A.

Subject index

- accession-specific 205
- Acer negundo* 130
- addition lines 164, 217
- alcohol dehydrogenase 2 166, 203
- alfalfa breeding 152
- alien addition lines 214
- alien chromosome 201, 205
- allele-specific PCR 5
- allotetraploid 246
- α -amylase 172, 263
- Arabidopsis* 310
- Arachis* 246
- Asimina triloba* 127, 130
- amphidiploidy 200, 214, 217
- ancestral chromosome 216
- Andean gene pool 262
- aneuploidy 40, 46, 126, 131, 132, 200, 330, 365
- anonymous cDNA clones 330
- anonymous genomic clones 201
- anonymous genomic DNA (gDNA) clones 330
- anther culture 164
- Avena atlantica* 360
- A. byzantina* 359
- A. sativa* (oat) 359
- A. hirtula* 360
- A. strigosa* 362
- A. wiestii* 362

- B genome maps 208
- basepair changes 2
- blackleg resistance 208
- Brassica* 126, 132
 - synteny maps 205
- Brassica carinata* 200
- B. insularis* 210, 217
- B. juncea* 200

- B. napus-nigra* 208
- B. napus* 200
- B. napus* map 213
- B. nigra* 200
- B. nigra* map 210
- B. oleracea* 200
- B. oleracea* maps 208
- B. rapa* 200
- B. rapa* maps 210
- B. rapa-oleracea* 205, 206, 210, 217
- breakage 207
- breeding system 128, 129
- Bremia lactucae* 226
- bronze 172
- bronze 1 2*
- bulked segregant analysis 5, 28, 232

- C genome maps 205
- catalogue of wheat genetic markers 328
- cDNA clones 202
- centromere 172, 180, 186, 189
- challenges 98
- challenges facing plant breeders 98
- chili pepper 131
- chlorophyll a/b binding protein 262
- chlorophyll binding protein 203
- chromosomal rearrangements 294
- chromosomal translocations 213
- chromosome evolution 131
- classification 124, 125
- clubroot resistance 202, 218
- clustering of loci 347
- colinear 295
- colinearity 294
- comparative genome mapping 131, 132, 133
- compensating nullisomic-tetrasomic lines 330
- Compositae 223

- conservation of gene order 354
 Core Map 240
 cowpea 135
 cruciferin 202
 cryptic structural differentiation 131
 cultivar development 152
Cucumis melo 2
 cytogenetic stocks 201
 cytogenetics 40, 46, 47, 145
 cytoplasmic introgression 125
 cytoplasmic male-sterility restorer 218
- degree of polymorphism 203
 deletions 207, 213, 214
 desaturation 208
 diploid 163
 disease resistance genes 224
 disomic addition lines 201
 disomic wheat-alien chromosome addition lines 330
 dispersed repetitive DNA 120
 ditelosomic lines 330, 347
 DNA content 216
 DNA isolation 10
 DNA sequence 4
 doubled haploid lines 19, 39, 44, 60
 Downy mildew 225
Drosophila melanogaster 2, 3
D. pseudoobscura 1, 3
D. simulans 3
 duplicated 294
 duplicated loci 132
 duplications 213, 214, 294
- early flowering 227
Eleusine coracana (finger millet) 294, 295
 elite germplasm 87
 embryogenesis abundant 202
 enhancement of heterosis 106
 epistasis 85
 erucic acid 208
 exotic germplasm 87
 evolution 116
- F-2 linkage maps 208
 fatty acid chain elongation 208
 favorable alleles 85
 female gametes 164
 fingerprinting 127, 205
 fingerprints 129
 footprint 3
- gene flow 133
 gene order conservation 131
 gene specific probes 202
 gene synteny relationships 354
 genetic distances 128
 genetic diversity 121, 123, 128, 129
 genetic linkage map 130
 genetic markers 201
 genetic or recombination lengths 168
 genetic variability 127, 128
 genome duplication 214, 304
 genome evolution 130
 genome relationships 200
 genome size 164
 genome specificity 202
 genome-specific 204
 genomic composition 213
 genomic formulae 216
 genotype by environment interaction 90
 germplasm sources 146
 glucosinolate 218
Gossypium hirsutum 129
 graphical genotype 52
 Gregor Mendel 261
- Helianthus* 133
 heterosis (or hybrid vigor) 99, 106
 heterozygosity 3
 homology 123, 126
Hordeum bulbosum 164
Hordeum vulgare (barley) 294, 295
 humans 2
 hybrid predictions 99
 hybrid species 131
 hybrid zones 133
 hypervariable regions (HVR) 121
- immortal 194
 immortalized F2, 241
 infraspecific diversity 126
 inhibitor 263
 insertion mutagenesis 31
 inter-genomic 214
 intergenomic homoeologous recombination 217
 International Triticeae Mapping Initiative 354
 interspecific aneuploids 217
 interstitial deletion 207
 interval mapping 58, 134
 least squares method 59
 maximum likelihood method 59
 intra-genomic 214
 intragenomic homoeologous 217
 intragenomic homoeologous recombination 216

- intragenomic recombination 217
 introgressant populations 131
 introgressed segments 108
 introgression 133, 134
 inversions 214
Iris 133
 irises 133
 isocitrate lyase 203
 isozyme 247
 isozyme allelic diversity 99
 isozyme allelic frequencies 102
 isozyme genotypes 100
 isozyme marker loci 102
 isozyme markers 292
 isozyme traits 292, 295
 isozymes 262, 263, 266
- known-function probes 330
- Lactuca* 126, 128
Lactuca sativa (lettuce) 223
 length polymorphisms 2
 lineage sorting 125
 linear least squares interval mapping (LIM)
 60
 linkage drag 85
 linkage integrity 101, 110
 linkage map construction 205
 linkage maps 130, 131
 linkage rearrangements 213
 LOD 51
Lycopersicon 125, 126, 128
Lycopersicon esculentum L. (tomato) 131,
 135, 310
- maintainer genes 218
 maize 2, 131, 132, 135, 240, 291, 292,
 294–296
 maize genome 240
 maize RFLP 295
 malate synthase 202
 male gametes 164
 map-based cloning 31
Mapmaker 262
 MAPMAKER-QTL 60
 mapping 146
 Hungarian Academy of Sciences, Szeged
 152
 Montana State University and University
 of Wisconsin 150
 University of Georgia 146
 population 16, 250
 marker allele frequency changes 102
 marker diversity 100
 marker-allele manipulation 103
 marker-assisted introgression 112
 marker-assisted selection (MAS) 85, 104,
 111
 marker-facilitated backcrossing 111, 112
 marker-facilitated introgression 109, 111
 marker-facilitated manipulation 103
 marker-facilitated selection 105
 Mesoamerican gene pool 262
 microsatellites 3, 4, 21, 255
 minisatellites 3, 21, 43
 molecular maps 240
 molecular marker analysis in alfalfa 156
 molecular markers 152
 morphological evolution 134
 morphological traits 134, 291, 292
 morphology 124
 multiple loci 84
 mungbean 135
 mutation rate 3
- napin 202
 naturally occurring polymorphism 2
 near-isogenic lines 28
 non-homologous recombination 217
 nuclear DNA markers 116, 117, 121–123
 null alleles 213
- oilseed *Brassicac*s 199
 origin of the genomes 215
Oryza sativa (rice) 287
 ortologous loci 215
 out-crossing 3
 overlapping translocations 214
- Panicum miliaceum* (proso millet) 294, 295
 paralogous loci 215
 PCR 46
 PCR 3, 4, 5
 peanut 246
 peanut breeding 255
Pennisetum glaucum (pearl millet) 294
 peroxidase 166
 phaseolin 262, 263
Phaseolus acutifolius 266
P. coccineus 261, 262, 266, 268
P. multiflorus 261
P. nanus 261
P. vulgaris 261, 266, 268
 phylogenetic analysis 123, 126
 phylogenetic inference 125
 phylogenetic relationships 205
 phylogeny 121, 124, 125
 physical lengths 168

- plant breeders 98
- plant improvement 97
- plasticity of the *Brassica* genomes 216
- plastocyanin 173
- polymerase chain reaction (PCR) 45
- polymorphism 1, 2, 3, 346
- polyploidy 131, 132, 327
- polysomic structure 214
- polysomy 216
- population size 3
- Populus* 134
- Populus tremuloides* 127
- positions of centromeres 347
- powdery mildew resistance 183
- predicting hybrid performance 99
- probe 4
- probe and map development 300
- progeny-specific maps 205
- pulsed field gel electrophoresis (PFGE) 47

- QTL 43, 44, 49, 54
- QTL 'alleles' 100
- quantitative trait loci (QTL) 32, 42, 134, 266, 296, 297, 302
- quantitative trait manipulation 103

- random amplified polymorphic DNA (RAPD) 4, 5, 18, 39, 42, 45, 46, 48, 50, 54, 119, 125, 247, 263, 296
- RAPD diversity 129
- RAPD markers 204
- rearrangements 194
- reciprocal 214
- recombinant inbred (RI) 39, 43, 241, 296
- recombinant inbred lines 17
- recombination 3
- recombinogenic 217
- resistance to blackleg disease 208, 218
- restriction fragment length polymorphism (RFLP) 10, 39, 40, 41, 43, 45, 48, 49, 52, 117, 118, 125, 127, 248, 291, 292, 294–297, 327
- RFLP analysis 118, 125
- RFLP diversity 118, 122, 128, 129
- RFLP genotypes 101
- RFLP loci 329, 330
- RFLP maps 240, 345
 - Arabidopsis thaliana* 159
 - barley 163
 - bread wheat 327
 - maize 240
 - peanut 246
 - potato 271
 - oats (*Avena* sp.) 359
 - rice 286
 - soybean 299
 - tomato 310
- RFLP markers 201
- RFLP nomenclature 328
- RFLP polymorphism 128
- RFLP-based genetic maps of the 21 chromosomes of wheat 350
- RFLP-based genetic maps of wheat chromosomes 3A, 3B, 3D, 7A, 7B, 7D and 4A 348
- ribosomal genes 203
- ribosomal RNA 187
- 2-row versus 6-row 173
- Rubus* 130

- satellites 167
- SCARs 230
- secondary polyploids 200, 213
- seed proteins 263
- seed storage proteins 266
- segregating populations 164
- segregation distortion 292, 295, 296
- selection gains 85
- self-incompatibility 202
- self-pollinating 3
- sequence databanks 4
- sequence duplication 213
- Setaria indica* (foxtail millet) 294, 295
- simple sequence repeats 3
- single strand conformation polymorphism 5
- sinigrin 208
- Solanum phureja* 272
- S. tuberosum* (potato) 131, 271
- Sorghum bicolor* (sorghum) 127, 129, 131, 132, 134, 291, 292, 294–297
- sorghum RFLP 295, 297
- Sorghum versicolor* 294
- Southern blotting 4
- soybean 4
- species-specific 205
- SSCP 25
- stability of C genome alien chromosomes 206
- structure of the *Brassica* genomes 213
- substitution 40, 46
- subtelomeric 173
- sugarcane 291, 294, 295
- syntenic differences 216
- synteny 5, 217
- synteny maps 205, 213
- systematics 116, 124

- tandem repeats 3

384 *Subject index*

- taxonomy 145
- telomeric 180, 189
- teosinte 135
- terminal 207
- translocations 167
- transposable elements 2, 194
- trisomics 167
- Triticum aestivum* (bread wheat) 327
- T. dicoccum* 346
- T. spelta* var. *duhamelianum* 346
- T. tauschii* 328

- validity of an interspecific genetic map 307
- variability 146

- variable number tandem repeat loci (VNTR)
121
- variable number tandem repeats 3
- Vigna* 135
- VNTR locus 21

- waxy* 2
- waxy* endosperm 168
- wheat relatives 353
- wheat RFLP mapping 328
- wheat-barley addition lines 168

- yeast artificial chromosomes 32

Advances in Cellular and Molecular Biology of Plants

1. R.L. Philips and I.K. Vasil (eds.): *DNA-Based Markers in Plants*. 1994
ISBN 0-7923-2714-4
2. E.G. Williams, A.E. Clarke and R.B. Knox (eds.): *Genetic Control of Self-Incompatibility and Reproductive Development in Flowering Plants*. 1994
ISBN 0-7923-2574-5